

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

ABSTRACT

Article type:

Research Article

Article history:

Received 29 December 2023
Received in revised form 30
January 2024
Accepted 4 February 2024
Published online

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 smart phones, 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 smart phones 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 (LightGBM). 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 LightGBM and XGB models achieved the best performance for the first and second dataset 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.

Keywords:

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

Cite this article: Shamohammadi, O; Pahlavani, P; & Sharifi, M. A. (2024). Estimating the travel modes using machine learning algorithms for sustainable urban transportation.



© The Author(s).

Publisher: University of Tehran Press.

DOI: <http://doi.org/00000000000000000000000000000000>

تخمین حالت های سفر به کمک الگوریتم های یادگیری ماشین به منظور حمل و نقل پایدار شهری

امید شامحمدی^۱ | پرهام پهلونی^۲ | محمدعلی شریفی^۳

۱. گروه GIS، دانشکده مهندسی نقشه برداری و اطلاعات مکانی، دانشکده گان فنی، دانشگاه تهران، تهران، ایران، رایانامه: omid.shamohammadi@ut.ac.ir

۲. نویسنده مسئول، گروه GIS، دانشکده مهندسی نقشه برداری و اطلاعات مکانی، دانشکده گان فنی، دانشگاه تهران، تهران، ایران رایانامه: pahlavani@ut.ac.ir

۳. گروه ژئودزی، دانشکده مهندسی نقشه برداری و اطلاعات مکانی، دانشکده گان فنی، دانشگاه تهران، تهران، ایران رایانامه: sharifi@ut.ac.ir

اطلاعات مقاله

چکیده

نوع مقاله:

مقاله پژوهشی

بخش زیادی از سفرهای روزانه درون شهری با هدف دسترسی به خدمات، سرویس ها و کالاهایی است که امکان تهیه آنها در یک منطقه مشخص وجود ندارد. از این رو، آنالیز مسیرهای پرتردد و علت حجم بالای ترافیک در آنها می تواند به توزیع صحیح تر امکانات، خدمات، و اختصاص کاربری مناسب به زمین ها با هدف کاهش تعداد، مسافت و زمان سفرهای درون شهری منجر شود. با ظهور سنسورهای تعیین موقعیت مبتنی بر سامانه های جهانی ناوبری ماهواره ای بر روی تلفن های همراه، امکان ثبت آنی مواردی چون موقعیت، سرعت و شتاب افراد به صورت آنی فراهم شده است. بنابراین، در این پژوهش تلاش شده است تا امکان استفاده از داده های GNSS ثبت شده توسط تلفن همراه، با هدف شناسایی حالت حمل و نقلی که کاربر از آن استفاده نموده است، توسط چهار مدل یادگیری ماشین نظارت شده با نام های *RF*، *GB*، *XGB* و *LightGM* مورد بررسی قرار گیرد. بدین منظور، دو مجموعه داده میکروسافت *Geolife* و *MTL 2017* که از ویژگی های لازم برای این هدف برخوردار هستند، به عنوان داده ورودی استفاده شده است. پس از استخراج ویژگی های هر مسیر از این دو مجموعه داده، با هدف بهبود کارایی مدل ها و کاهش حجم پردازش، از میان ویژگی های موجود، ویژگی های مهم تر شناسایی و کلاسه بندی بر اساس آنها اعمال شده است. در بین مدل های استفاده شده، مدل *LightGM* برای مجموعه داده ای اول با کسب مقدار ۹۲/۵۷ درصد در *FI-Score* برای داده های تست و مدل *XGB* برای مجموعه داده ای دوم با کسب مقدار ۹۲/۶۷ در *FI-Score* برای داده های تست در در بین مدل های موجود بهترین عملکرد را داشته اند. از تعداد ۱۳۴۹ سفر موجود، ۱۲۵۰ سفر توسط این الگوریتم به درستی تخمین زده شده است که منجر به حمل و نقل پایدار شهری می گردد.

کلیدواژه ها: حمل و نقل پایدار شهری، حالت های سفر، جریان داده های GNSS، استخراج ویژگی، الگوریتم های کلاسه بندی نظارت شده.

تاریخ دریافت: ۱۴۰۲/۱۰/۰۹

تاریخ بازنگری: ۱۴۰۲/۱۱/۱۰

تاریخ پذیرش: ۱۴۰۲/۱۱/۱۵

تاریخ انتشار:

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



© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <http://doi.org/00000000000000000000000000000000>

مقدمه و بیان مسأله

انسان همواره برای رفع نیازهای خود، زنده ماندن و تجسس و تفتیش در محل زندگی، از ابتدای خلقت تاکنون در حال جابه‌جایی بوده است (Schlebusch & Jakobsson, 2018). گسترش محیط‌های شهری و افزایش بی‌رویه جمعیت در کلان‌شهرها، سبب افزایش تقاضا برای جابه‌جایی‌های درون شهری شده است که از اثرات آن می‌توان به ایجاد ترافیک، آلودگی هوا، هدر رفت زمان و انرژی اشاره نمود. از این‌رو، ارزیابی دقیق رفتار حرکتی افراد در جامعه، به منظور درک حرکت و تأثیر آن بر فعالیت‌های شهروندان و حفاظت از محیط‌زیست امری ضروری برای برنامه‌ریزان شهری در راستای حمل و نقل پایدار شهری می‌باشد (Martín-Baos et al., 2023). حمل و نقل دومین هزینه کلان در زندگی افراد است به گونه‌ای که در مطالعات اتحادیه اروپا و آمریکا، شهروندان به ترتیب به طور متوسط ۱۵ و ۲۵ درصد از کل هزینه‌های زندگی خود را صرف حمل و نقل می‌کنند (Wang et al., 2018)، این مقدار می‌تواند به عنوان معیاری از اهمیت حمل و نقل در فعالیت‌های انسانی در نظر گرفته شود.

بخش بسیاری از سفرهای صورت گرفته در طول روز با هدف دسترسی به کالاها، امکانات و خدماتی است که افراد در نزدیکی محل سکونت خود به آن‌ها دسترسی ندارند به عنوان مثال 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 حمل و نقل گردد (Sun et al., 2023). آنالیز جابجایی‌های صورت گرفته در سطح شهر و همچنین فرم حمل و نقل مورد استفاده در هر یک از آن‌ها علاوه بر آنکه می‌تواند درک مناسبی از مسیرهای پرتردد فراهم آورد، در شناسایی هدف جابجایی صورت گرفته نیز تاثیرگذار است (Kashifi et al., 2022). از این رو مدیران شهری می‌توانند، متناسب با اهداف جابجایی‌ها، به بهبود توزیع امکانات در سطح شهر با ابزار تخصیص کاربری زمین پردازند. چنین رخدادی باعث کاهش حجم سفرهای روزانه، مصرف انرژی، تولید گازهای گلخانه‌ای و صرفه‌جویی در زمان می‌گردد.

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

با توجه به توانایی الگوریتم‌های یادگیری ماشین به عنوان ابزاری مناسب برای طبقه‌بندی اطلاعات گوناگون، این پژوهش از چهار مدل یادگیری ماشین برای شناسایی حالت سفر استفاده کرده‌است. برای این منظور، مجموعه‌ای از ویژگی‌ها از داده‌های ثبت شده توسط سنسور GNSS تلفن‌های همراه کاربران استخراج شده‌است. با هدف بهبود عملکرد و کاهش حجم پردازش‌های صورت گرفته شناسایی ویژگی‌های موثر 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 شده‌است.

پیشینه پژوهش

با پیشرفت و هوشمندی تلفن‌های همراه، تعداد حسگرهای مورد استفاده در ساخت آن‌ها رو به افزایش است. مطالعات زیادی از داده‌ها و حسگرهای متفاوتی برگرفته از داده‌های ثبت شده توسط تلفن‌های همراه هوشمند برای شناسایی حالت‌های حمل‌ونقلی استفاده کرده‌اند. برخی از مطالعات پیشین چالش استفاده تنها از داده‌های GPS¹ به صورت خام، بدون استفاده از داده‌های مکمل و حسگرهای دیگر را مورد بررسی و پیش‌بینی قرار داده‌اند (Nawaz et al., 2020). به عنوان مثال در (Li et al., 2020) از مجموعه داده‌های GPS ثبت شده در شهر پکن کشور چین، در (Song et al., 2016) از داده‌های GPS خام ثبت شده در کشور ژاپن، در (Yazdizadeh et al., 2021) از داده‌های ثبت شده در شهر مونترال کشور کانادا و در (Bolbol et al., 2012) از داده‌های شهر لندن کشور انگلیس استفاده شده است.

در برخی از مطالعات، از حسگرهای دیگر بدون استفاده و وابستگی به حسگر GPS استفاده شده است. به عنوان مثال در (Friedrich et al., 2020) از مجموعه داده‌های مرکز تحقیقاتی SHL² شامل حسگرهای شتاب‌سنج،ژیروسکوپ، شتاب خطی، مغناطیس‌سنج، فشارسنج، جهت‌سنج، گرانش‌سنج و موارد دیگر استفاده شده‌است.

مطالعات دیگری از حسگرهای هوشمند به عنوان حسگرهای مکمل در کنار داده‌های اخذ شده از حسگر GPS استفاده می‌کنند. در (Bantis & Haworth, 2017) از حسگر شتاب‌سنج در کنار حسگر GPS در (Byon & Liang, 2014) از حسگرهای شتاب‌سنج، مغناطیس‌سنج و GPS و در (Jahangiri & Rakha, 2015) از حسگرهای شتاب‌سنج،ژیروسکوپ و بردار چرخش در کنار حسگر GPS استفاده شده است.

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

الگوریتم‌های متفاوتی به جهت کلاسه‌بندی حالت‌های حمل‌ونقلی در مطالعات استفاده شده است که به‌طور کلی می‌توان 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 از پیش تعیین شده و به صورت مرحله‌ای به کلاسه‌بندی می‌پردازد (Gong et al., 2012). به عنوان مثال در (Bohte & Maat, 2009) اگر میانگین سرعت کمتر از ۲۵ کیلومتر بر ساعت و بیشینه سرعت کمتر از ۴۵ کیلومتر بر ساعت باشد حالت حمل‌ونقلی دوچرخه‌سواری پیش‌بینی می‌شود. در برخی دیگر از مطالعات منطق فازی پیاده‌سازی شده، در این مطالعات با تعریف توابع عضویت، تمایز بین حالت‌های حمل‌ونقلی مختلف صورت گرفته به گونه‌ای که احتمال رخداد یک حالت حمل و نقلی خاص بین ۰ و ۱ تنظیم شده است (Sauerländer-Biebl et al., 2017).

1 Global Positioning System

2Sussex-Huawei Locomotion-Transportation

امروزه روش های یادگیری ماشین و یادگیری عمیق بیش از باقی الگوریتم ها مورد استفاده قرار می گیرد، پیش بینی حالت های حمل و نقلی نیز از این مورد پشتیبانی می کند. چندین نوع الگوریتم در مطالعات بررسی شده و مورد استفاده و آزمایش قرار گرفته اند، در برخی از مطالعات (Nawaz et al., 2020) و (Yazdizadeh et al., 2019) الگوریتم های یادگیری ماشین مدل های RF و در (Li et al., 2021) و

Wang et al., (2018) مدل های GB استفاده شده است.

به عنوان یک جمع بندی، استفاده از روش های مبتنی بر قانون و تئوری فازی نیازمند تعیین قوانین و توابع عضویت به تفکیک هریک از ویژگی های استخراجی می باشد. این مورد در صورت افزایش تعداد ویژگی ها با مشکل روبرو می شود، همین طور در روش های یادگیری عمیق ویژگی ها در لایه موسوم به لایه پنهان استخراج خواهد شد و مهندسی ویژگی نقش کلیدی در پیش بینی خواهد داشت به گونه ای که در بیشتر موارد موجب کاهش دقت در مقایسه با استخراج ویژگی دستی می شود. همچنین اگرچه در تحقیقات پیشین، طبقه بندی حالت های حمل و نقل بررسی شده است، در این تحقیق تلاش شده است تا مقایسه ای از عملکرد چندین روش طبقه بندی مطرح همراه با انتخاب ویژگی های موثر صورت پذیرد. به منظور بهبود جامعیت مقایسه صورت گرفته، از تعداد ویژگی های بیشتری استفاده شده است تا ارزیابی دقیق تری بر روی عملکرد الگوریتم های انتخاب ویژگی در زمانی که کاهش ابعاد بیشتری مد نظر است صورت پذیرد.

محدوده و قلمروی مورد مطالعه

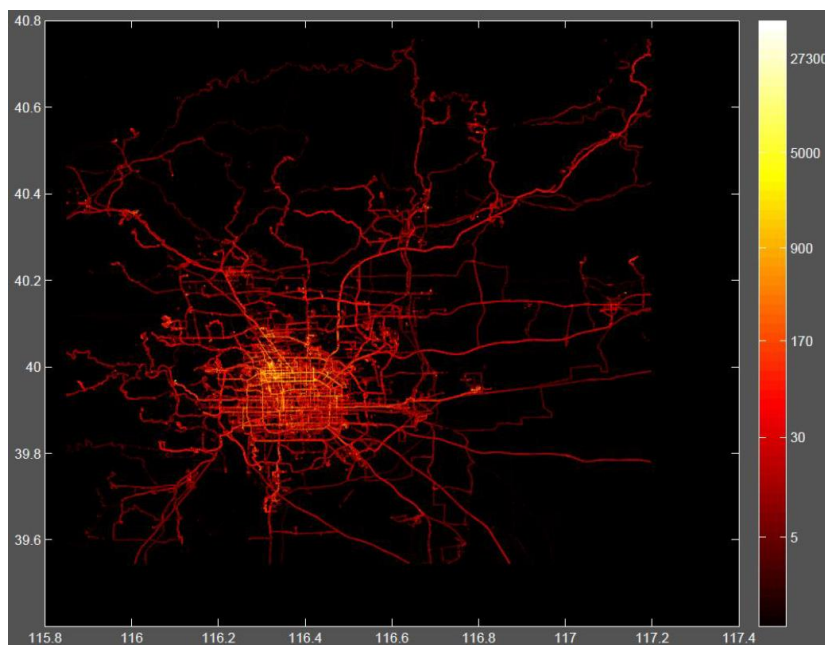
در این تحقیق برای ارزیابی روش پیشنهادی از مجموعه داده های پروژه Geolife که توسط مرکز تحقیقاتی مایکروسافت تهیه و ذخیره شده است، استفاده شده است. این مجموعه داده طیف وسیعی از رفت و آمدهای کاربران را با مقاصد مختلف از جمله مکان های کاری، استادیوم ها، مراکز خرید و سرگرمی را شامل می شود. ذخیره سازی این داده ها به صورت زمانی - مکانی در مدت ۵ سال (آوریل ۲۰۰۷ تا آگوست ۲۰۱۲) و توسط ۱۸۲ نفر انجام پذیرفته است که ۶۹ کاربر در این مجموعه داده، سفرهای خود را برچسب گذاری کرده اند. اطلاعات مربوط به این مجموعه داده در جدول ۱ نشان داده شده است.

جدول ۱. مشخصات سفرهای موجود در داده Geolife

تعداد کل کاربران	تعداد کل سفرهای انجام گرفته	کل مسافت های پیموده شده (بر حسب کیلومتر)	کل زمان های پیموده شده (حسب ساعت)
۱۸۲	۱۷۶۲۱	۱۲۹۲۹۵۱۱	۵۰۱۷۶

این مجموعه داده به صورت منبع باز و رایگان در صفحه سازمان مایکروسافت قابل بارگیری و استفاده می باشد، Geolife Dataset, (Accessed June 2022). خط سیر ثبت شده در این مجموعه داده توسط دستگاه های ثبت موقعیت مکانی PDA و تلفن همراه هوشمند با نرخ نمونه برداری متفاوت ثبت و ذخیره شده اند. این مجموعه داده در بیش از ۳۰ شهر در کشورهای مختلف مانند چین، ایالات متحده آمریکا و برخی از کشورهای اروپایی جمع آوری شده است. شکل ۱ توزیع مجموعه داده ها را در پکن نشان می دهد، شهری که بیشتر داده ها در آن جمع آوری شده است.

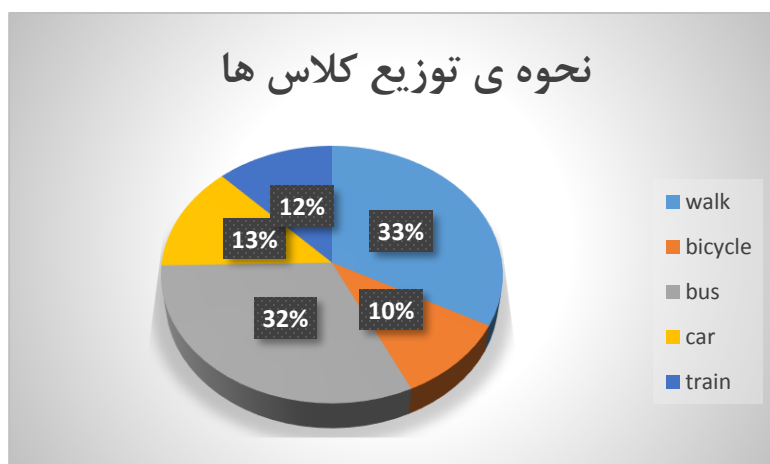
در این مجموعه داده حالت های سفری که برچسب گذاری شده اند شامل: پیاده روی، دوچرخه، اتومبیل، اتوبوس، قطار، مترو، هواپیما، دویدن، موتورسیکلت، تاکسی و جابه جایی با استفاده از قایق می باشد.



شکل ۱. توزیع داده‌ها در شهر پکن

در این تحقیق، از حالت‌های حمل‌ونقل غیرزمینی اعم از هواپیما و قایق صرف نظر شده است. در بین حالت‌های زمینی نیز با توجه به کافی نبودن داده‌ها در حالت‌های دویدن و موتورسیکلت این دو حالت نیز کنار گذاشته شده‌اند. همچنین با توجه به فایل راهنمای موجود خود مایکروسافت برای داده‌های Geolife، دو حالت استفاده از اتومبیل و استفاده از تاکسی، هر دو به‌عنوان یک حالت کلی استفاده از اتومبیل در نظر گرفته شده‌است. در ادامه، با توجه به اینکه در شهر پکن مسیر قطارهای شهری و مترو به هم متصل و یکپارچه می‌باشند، از طرفی همچون استاندارد خاصی برای تفکیک این دو وجود ندارد شاید در برچسب‌گذاری اشتباه کنند، از این رو این دو حالت نیز به‌عنوان حالت حمل‌ونقل استفاده از قطار در نظر گرفته شده‌است. در نهایت با حذف و ادغام حالت‌های گفته شده پنج حالت پیاده‌روی، دوچرخه، اتوبوس، ماشین و قطار به‌عنوان کلاس‌های نهایی در این مطالعه مورد بررسی قرار گرفته‌شد. در شکل ۲ نحوه‌ی توزیع داده‌ها را در هر یک از حالت‌ها به نمایش گذاشته شده‌است.

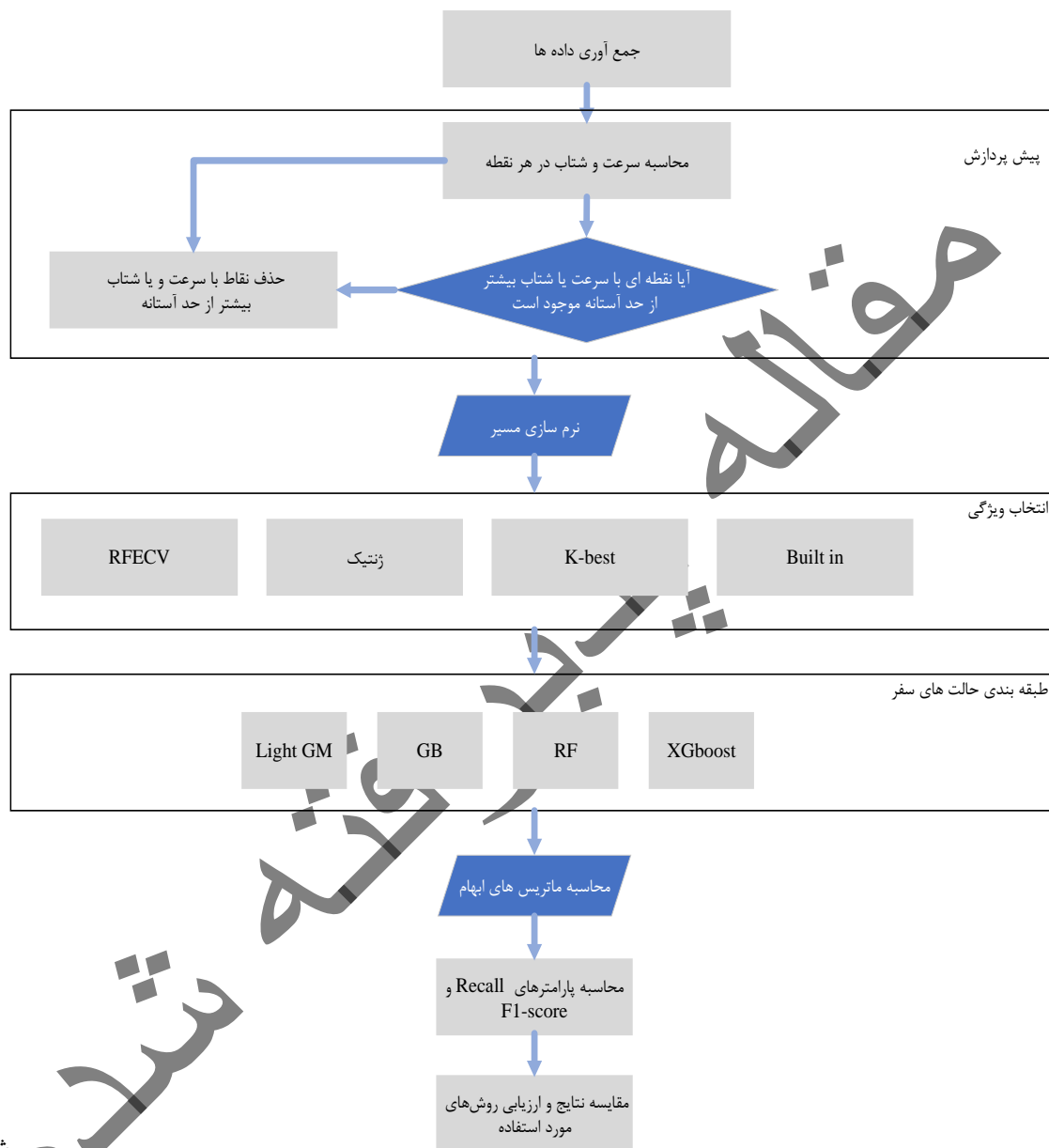
شکل ۲



شکل ۲. نحوه‌ی توزیع کلاس‌ها

روش پژوهش

در شکل ۳ فلوچارت روش تحقیق برای پیش‌بینی و شناسایی حالت‌های حمل‌ونقل نشان‌داده شده است.

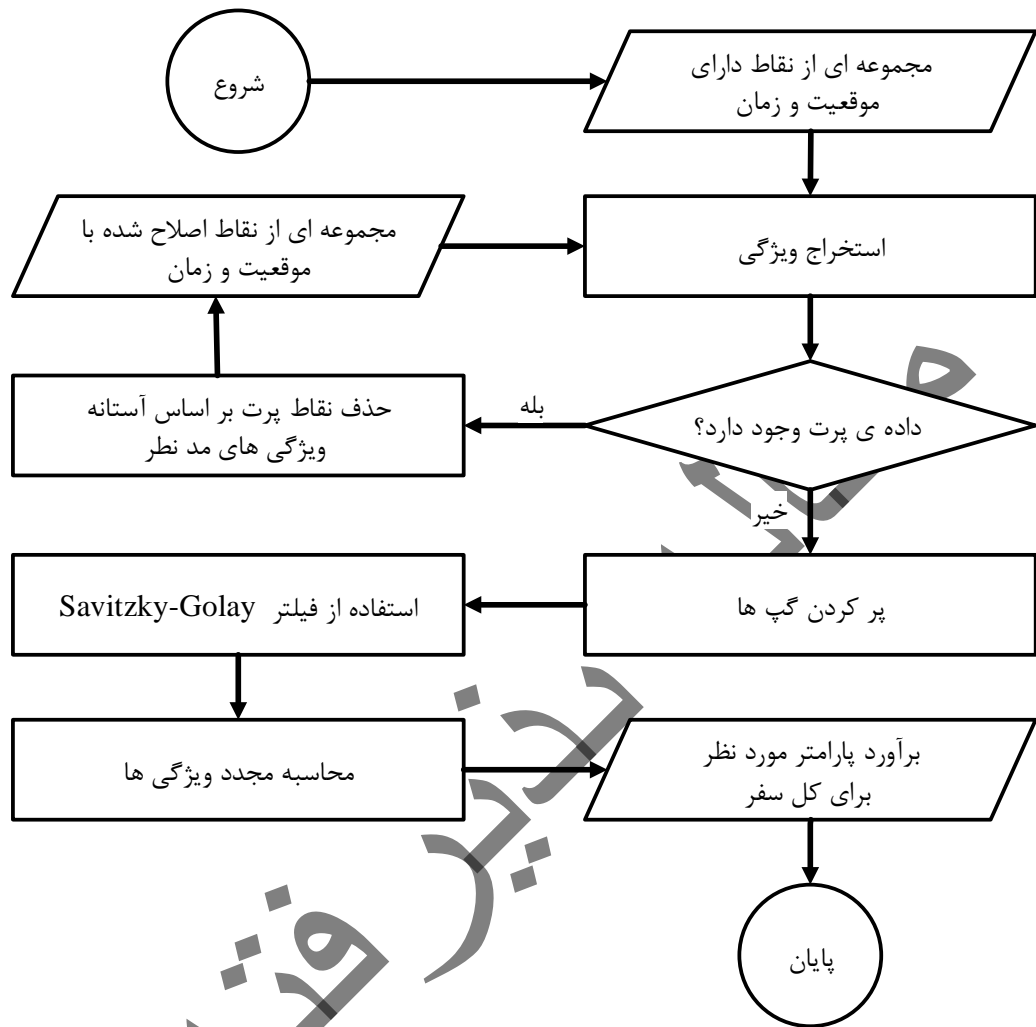


شکل ۳.

فلوچارت مربوط به روش تحقیق

تجزیه و تحلیل داده‌ها

برای انجام پیش‌پردازش داده‌ها به ترتیب مراحل زیر صورت گرفته است:



شکل ۴. فلوچارت مربوط به انجام پیش پردازش ها

در این مرحله ابتدا با استفاده از نقاط خام موجود در داده ها برای هر کدام از نقاط موجود سرعت و شتاب محاسبه می شود، سپس با استفاده از حد آستانه تعریف شده در جدول ۲ نقاطی که از این حد آستانه دارای سرعت و شتاب بیشتری هستند از مجموعه داده ها حذف شدند.

جدول ۲. مقادیر حد آستانه برای شتاب و سرعت

حالت های حمل و نقلی	حد آستانه سرعت (m/s)	حد آستانه شتاب (m/s ²)
پیاده روی	۷	۳
دوچرخه سواری	۱۲	۳
اتوبوس	۳۴	۲
اتومبیل	۵۰	۱۰
قطار	۳۴	۳

سپس نقاطی که دیگر سرعت و شتاب آن‌ها از حد آستانه بیشتر نیست مورد بررسی قرار می‌گیرند که آیا از نظر زمانی دارای گپ هستند یا خیر. در هر سفر این پارامتر بررسی می‌شود و اگر بین دو نقطه بیش از مدت‌زمان مشخص شده گذشته باشد با استفاده از تابع درون‌یابی گپ تشخیص داده شده بین دو نقطه پر شده است.

در مرحله‌ی بعدی تمام مسیرهای موجود با استفاده از فیلتر Savitzky-golay نرم شده‌اند. پس از این مرحله مجدداً تمام ویژگی‌های تعریف شده برای تمام مسیرهای موجود محاسبه شده‌اند. در نهایت یک فایل اکسل با ۱۳۴۹ رکورد ذخیره شده‌است که هر رکورد نشان‌دهنده‌ی یک سفر مربوط به کاربر با شماره‌ی موردنظر می‌باشد.

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

جدول ۳. ویژگی‌های استخراج شده برای هر سفر

فیلتر
فنا
شده

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	VelChangePar
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	accChangePar	۳۹	jerkSkew
۱۴	jerkMean(m/s ³)	۴۰	jerkKurtosis
۱۵	jerkStd(m/s ³)	۴۱	jerkIqr
۱۶	jerkChangePar	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)

مدل های کلاسه بندی

حال پس از پیش برداشته شدن موارد موردمشاهده و همچنین استخراج ویژگی‌های موردنیاز، پیاده‌سازی مدل‌های کلاسه‌بندی صورت پذیرفته است. برای پیاده‌سازی مدل‌های نامبرده شده از کتابخانه‌های Scikit-learn, XGBoost و LightGM موجود و با بهره‌گیری از زبان برنامه‌نویسی پایتون پیاده‌سازی شده‌اند. به این منظور برای هر کدام از مدل‌های موردنظر حالت‌های زیر پیاده‌سازی شده‌است:

پیاده‌سازی مدل با استفاده از پارامترهای پیش‌فرض.

بهینه‌کردن ابر پارامترها برای هر مدل و پیاده‌سازی مجدد.

استخراج ویژگی‌های مؤثر با استفاده از Kbest و پیاده‌سازی مجدد مدل

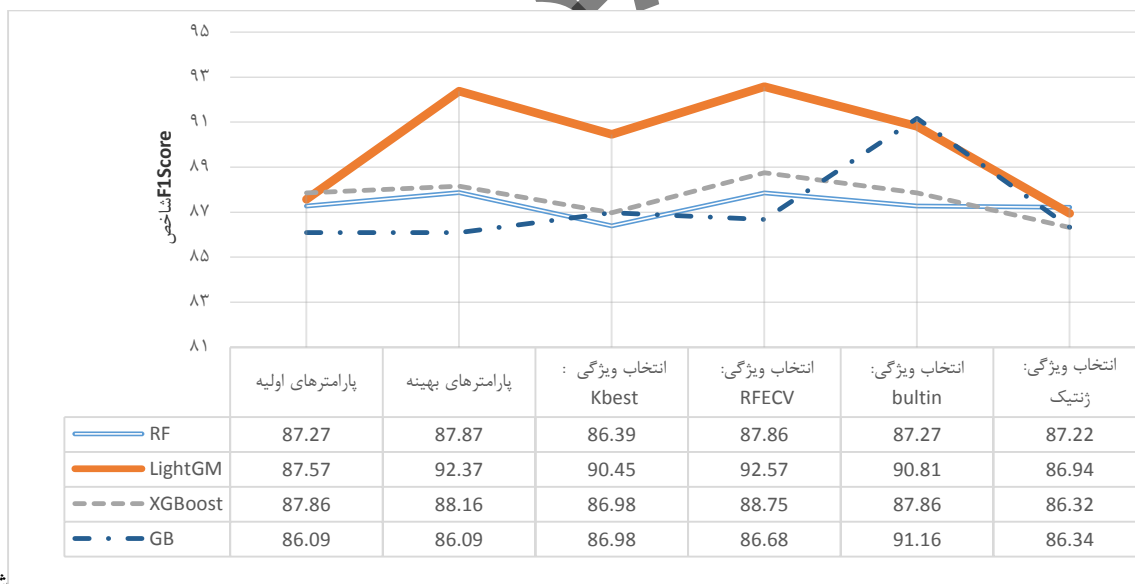
استخراج ویژگی‌های مؤثر با استفاده از RFECV و پیاده‌سازی مجدد مدل

استخراج ویژگی‌های مؤثر با استفاده از Built-in و پیاده‌سازی مجدد مدل

استخراج ویژگی‌های مؤثر با استفاده از الگوریتم ژنتیک و پیاده‌سازی مجدد مدل

یافته‌های پژوهش

به منظور بررسی و ارزیابی نتایج به دست آمده برای مدل‌های کلاسه‌بندی از پارامترهای F1-Score، بازخوانی و دقت استفاده شده است. از آنجایی که برجسته‌ترین 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 کدام از پارامترهای گفته شده در قالب شکل‌های ۵ تا ۷ به نمایش گذاشته شده‌است.

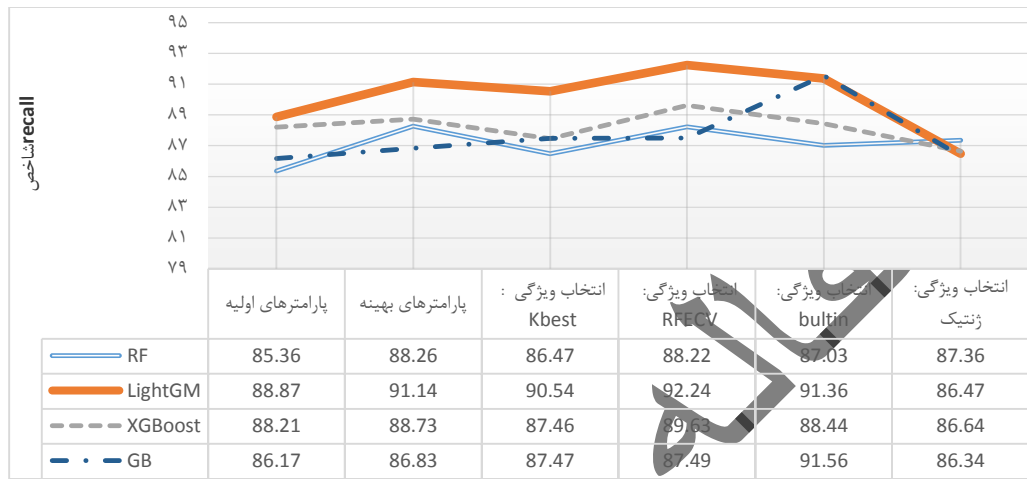


شکل ۵

میانگین وزنی F1-Score در مدل‌های کلاسه‌بندی

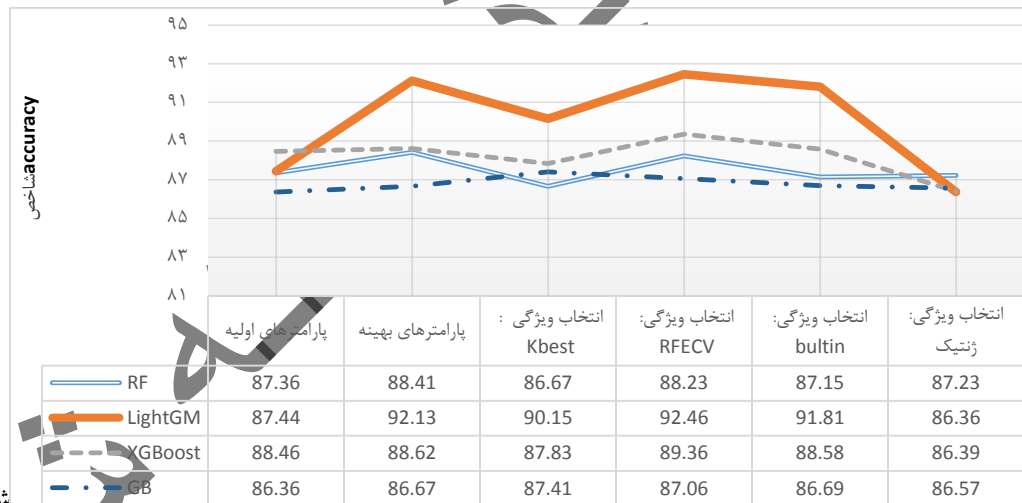
- 4 Recall
- 5 Accuracy

همانطور که از شکل ۵ مشخص است دقت مدل‌های به کار گرفته شده بسیار نزدیک بهم می‌باشد ولی مدل LightGM توانسته در بحث استخراج ویژگی‌های موثر نسبت به سایر مدل‌ها کارایی بیشتری از خود نشان دهد و با کسب ۹۲/۵۷ درصد در بین داده‌های تست در حالت انتخاب ویژگی با مدل RFECV بالاترین مقدار را در این پژوهش کسب کند.



شکل ۶. شاخص

Recall در مدل‌های کلاسه‌بندی



شکل ۷. شاخص

accuracy در مدل‌های کلاسه‌بندی

این دو شاخص نیز نتایج مشابهی هم چون شاخص F1-Score ارائه می‌دهند، مدل LightGM به عنوان مدل برتر و پس از آن مدل XGBoost بوده است 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 کدام از کلاس‌ها شاخص F1-Score را به تفکیک مدل و حالت انتخابی برای هر کدام از کلاس‌ها به همراه تعداد ویژگی انتخابی در قسمت انتخاب ویژگی در جدول ۴ نشان داده شده است.

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

مدل کلاسه‌بندی	حالت انتخابی	پیاده	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 شده‌اند که این امر می‌تواند ناشی از تمایز سرعت پایین فرد در حالت پیاده با سایر حالت‌ها و در حالت قطار بخاطر مستقیمی مسیر طی شده باشد. در ادامه برای مقایسه بهتر کارایی هر کدام از مدل‌های به کار گرفته شده، در جدول ۵ تعداد کلاس‌های تخمین شده برای هر کدام از کلاس‌های موجود آورده شده‌است.

جدول ۵. بررسی تعداد سفرهای تخمین شده توسط هر کدام 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
	GA	417	108	367	132	149	1173
LightGBost	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
	Built-in	416	118	397	148	154	1233
	GA	416	105	363	128	148	1160
XGBoost	پارامترهای اولیه	416	111	371	135	149	1182
	پارامترهای بهینه	421	115	367	132	153	1188
	Kbest	421	110	363	128	151	1173
	RFECV	420	112	371	139	154	1196
	Built-in	421	115	371	130	149	1186
	GA	420	108	363	123	143	1157

1159	154	123	354	111	417	پارامترهای اولیه	GB
1159	154	123	354	111	417	پارامترهای بهینه	
1168	156	126	359	107	420	Kbest	
1169	151	125	363	110	420	RFECV	
1170	153	128	359	114	416	Built-in	
1152	146	119	359	107	421	GA	

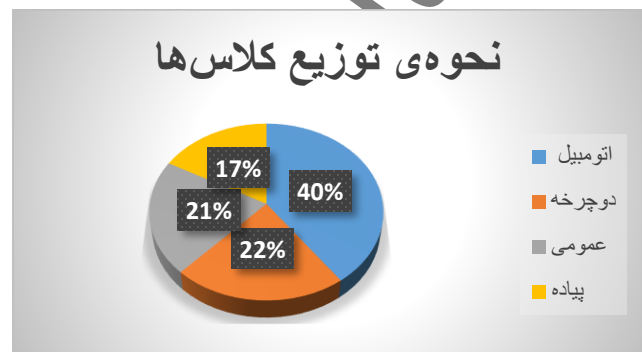
در پایان ارزیابی، دقت حاصل از روش تحقیق این مطالعه با سایر تحقیقات انجام شده مورد مقایسه قرار گرفته است. در این قسمت هشت مطالعه اخیر که بروی داده‌ی Geolife صورت گرفته‌اند به همراه حالت‌های حمل‌ونقلی که مورداستفاده قرار گرفته است با روش پیشنهادی این مطالعه مقایسه گردیده‌اند که نتایج حاصل از این مقایسه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

جدول ۶. مقایسه F1-Score به دست آمده با تحقیقات اخیر

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

در ادامه برای ارزیابی روش پیشنهادی و کارایی آن لازم است تا بروی یک مجموعه داده‌ی دیگر نیز پیاده‌سازی صورت گرفته‌شود از این رو در ابتدا مجموعه داده‌ی دوم به صورت مختصر معرفی و سپس نتایج به دست آمده نیز به نمایش گذاشته می‌شود.

این مجموعه داده در تاریخ ۲۰۱۷/۰۹/۱۸ تا ۲۰۱۷/۰۱/۱۸ در کشور کانادا، شهر مونترال، بزرگترین شهر استان کبک، توسط آزمایشگاه تحقیقات برنامه حمل و نقل و همکاری دانشگاه کنکور دیا ثبت و ذخیره شده است. جمع آوری داده‌ها از طریق یک برنامه قابل نصب بر روی تلفن همراه هوشمند (در هر دو سیستم عامل IOS و Android) انجام شده است که به طور خودکار خط سیر کاربر را ثبت می‌کند و هنگامی که کاربر در یک مکان معین به مدت بیش از ۱۲۰ ثانیه متوقف شوند، از کاربر می‌خواهد تا اطلاعات سفر خود را تکمیل کنند. این اطلاعات شامل برچسب‌دهی حالت سفر و هدف از سفر می‌باشد. حالت‌های حمل و نقلی پس از ترجمه شامل پیاده‌روی، استفاده از اتوبوس/موتورسیکلت، استفاده از تاکسی، استفاده از حمل و نقل عمومی، استفاده از دوچرخه و استفاده از اتوبیل‌های اجاره‌ای می‌باشد. در این مطالعه حالت‌های حمل و نقلی استفاده از تاکسی، استفاده از اتوبوس/موتورسیکلت و استفاده از اتوبیل‌های اجاره‌ای به عنوان حالت استفاده از اتوبیل در نظر گرفته شده‌اند. شکل نحوه توزیع حالت‌های حمل و نقلی مورد استفاده در این مجموعه داده را نمایش می‌دهد.



شکل ۸. نحوه‌ی توزیع کلاس‌ها در مجموعه داده دوم

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

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

تعداد ویژگی	Weighted	عمومی	اتوبیل	دوچرخه	پیاده	حالت انتخابی	مدل کلاسه‌بندی
52	87.86	61.7	91.58	95.66	87.92	پارامترهای اولیه	Random Forest

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

نتیجه و پیشنهاد

با شناسایی دقیق حالت حمل و نقل هر کاربر، می توان درک واقعی تری از تعداد وسایل نقلیه مورد نیاز برای جابه جایی 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 می‌توان با ایجاد خطوط پرسرشتین (HOV) در اتوبان‌های شهری تمایل استفاده به اتومبیل‌های تک‌سرنشتین را کاهش داد.

چالش‌های متنوعی تا رسیدن به هدف پژوهش وجود دارد. امروزه به لطف حسگرهای موجود در تلفن‌های هوشمند، داده‌های متنوعی توانایی اخذ و ذخیره شدن را دارند. چالش اول انتخاب داده‌های مورد استفاده در پیش‌بینی می‌باشد به گونه‌ای که علاوه بر دردسترس بودن، بالاترین سرعت و کمترین هزینه را در برداشته باشد. چالش دوم وجود داده‌های پرت و خطا به دلیل وجود خطاهای سیستمی و تصادفی در زمان اخذ داده‌ها می‌باشد که می‌تواند دقت را به صورت مستقیم تحت شعاع خود قرار دهد. چالش بعدی استخراج ویژگی از مجموعه داده‌ها است، استخراج ویژگی یکی از مهم‌ترین مراحل در پیش‌بینی می‌باشد 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 می‌باشد به گونه‌ای که موجب کاهش هزینه پردازشی در نتیجه افزایش سرعت و بهبود دقت در پیش‌بینی گردد. چالش نهایی پیاده‌سازی و انتخاب یک الگوریتم کلاسه‌بندی نظارت شده می‌باشد که علاوه بر قدرتمند بودن، بتوان با سرعت بالاتری نتیجه را پیش‌بینی کرد.

اگرچه در تحقیقات پیشین تمرکز بیشتر مطالعات تنها بر استخراج ویژگی، انتخاب ویژگی‌ها و کلاسه‌بندی بوده است، در این تحقیق تلاش شده است تا مقایسه‌ای بر روی عملکرد روشهای مختلف در انجام این مهم فراهم آورده شود. 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 حمل و نقل استفاده شده‌اند. به منظور بهبود جامعیت مقایسه صورت گرفته، تعداد ویژگی‌های بیشتری نیز استفاده شد تا ارزیابی دقیق‌تری بر روی عملکرد الگوریتم‌های انتخاب ویژگی در زمانی که کاهش ابعاد بیشتری مد نظر بود صورت پذیرد.

نتایج بدست آمده در این پژوهش از دقت مطلوبی در برآورده ساختن اهداف و طبقه‌بندی حالت‌های سفر در راستای حمل و نقل پایدار شهری برخوردار است. با این وجود، 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 حمل و نقل 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 حمل و نقلی با استفاده از تلفن‌های هوشمند را جایگزینی بهتر از پرسش‌نامه‌ها در نظر گرفت که با هزینه پایین و سرعت بالایی حجم عظیمی از داده‌ها را ثبت و ذخیره می‌کند.

با توجه به اثبات کارآمدی روش ارائه شده پیشنهاد می‌شود در ادامه تحقیق و در راستای جمع‌آوری داده‌های بومی هر شهر در کشور با مراجعه و هماهنگی با سازمانهای ذیربط، برنامه‌ای در سیستم عامل اندروید تهیه 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 بازه‌ی زمانی موردنظر نماید و در انتهای سفر هم با پایان یافتن سفر برچسب حالت جابه‌جایی را از کاربر گرفته و ذخیره کند.

همچنین چون مجموعه داده‌ی مورد استفاده دارای تعدادی زیادی از کاربران می‌باشد که سفرهای خود را برچسب‌گذاری نکرده‌اند در ادامه تحقیق توانایی روش‌های نیمه نظارت شده مورد ارزیابی قرار گیرد و دقت‌های به دست آمده از این روش برای داده‌های دارای برچسب‌گذاری شده با روش نظارت شده استفاده شده در این تحقیق مقایسه گردد.

References

- "Geolife Dataset." <https://www.microsoft.com/en-us/research/publication/geolife-gps-trajectory-dataset-user-guide> (accessed).
- A. Bolbol, T. Cheng, I. Tsapakis, and J. Haworth, "Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification," *Computers, Environment and Urban Systems*, vol. 36, no. 6, pp. 526-537, 2012.
- A. Jahangiri and H. A. Rakha, "Applying machine learning techniques to transportation mode recognition using mobile phone sensor data," *IEEE transactions on intelligent transportation systems*, vol. 16, no. 5, pp. 2406-2417, 2015.
- A. Nawaz et al., "Mode Inference using enhanced Segmentation and Pre-processing on raw Global Positioning System data," *Measurement and Control*, vol. 53, no. 7-8, pp. 1144-1158, 2020.
- A. Nawaz, H. Zhiqiu, W. Senzhang, Y. Hussain, I. Khan, and Z. Khan, "Convolutional LSTM based transportation mode learning from raw GPS trajectories," *IET Intelligent Transport Systems*, vol. 14, no. 6, pp. 570-577, 2020.
- A. Sauerländer -Biebl, E. Brockfeld, D. Suske, and E. Melde, "Evaluation of a transport mode detection using fuzzy rules," *Transportation research procedia*, vol. 25, pp. 591-602, 2017.
- A. Yazdizadeh, Z. Patterson, and B. Farooq, "An automated approach from GPS traces to complete trip information," *International Journal of Transportation Science and Technology*, vol. 8, no. 1, pp. 82-100, 2019.
- A. Yazdizadeh, Z. Patterson, and B. Farooq, "Semi-supervised GANs to infer travel modes in GPS trajectories," *Journal of Big Data Analytics in Transportation*, vol. 3, no. 3, pp. 201-211, 2021.
- 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).
- B. Friedrich, C. Lübbe, and A. Hein, "Analyzing the importance of sensors for mode of transportation classification," *Sensors*, vol. 21, no. 1, p. 176, 2020.
- B. Wang, Y. Wang, K. Qin, and Q. Xia, "Detecting transportation modes based on LightGBM classifier from GPS trajectory data," in *2018 26th International Conference on Geoinformatics, 2018: IEEE*, pp. 1-7.
- C. M. Schlebusch and M. Jakobsson, "Tales of human migration, admixture, and selection in Africa," *Annual Review of Genomics and Human Genetics*, vol. 19, pp. 405-428, 2018.
- 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.
- H. Gong, C. Chen, E. Bialostozky, and C. T. Lawson, "A GPS/GIS method for travel mode detection in New York City," *Computers, Environment and Urban Systems*, vol. 36, no. 2, pp. 131-139, 2012.
- J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189-1232, 2001.
- J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques Third Edition [M]," *The Morgan Kaufmann Series in Data Management Systems*, vol. 5, no. 4, pp. 364-371, 2012.
- J. Li, X. Pei, X. Wang, D. Yao, Y. Zhang, and Y. Yue, "Transportation mode identification with GPS trajectory data and GIS information," *Tsinghua Science and Technology*, vol. 26, no. 4, pp. 403-416, 2021.
- Kashifi, M. T., Jamal, A., Kashifi, 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.

- L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5-32, 2001.
- L. Li, J. Zhu, H. Zhang, H. Tan, B. Du, and B. Ran, "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, pp. 282-292, 2020.
- 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*.
- S. Dabiri and K. Heaslip, "Inferring transportation modes from GPS trajectories using a convolutional neural network," *Transportation research part C: emerging technologies*, vol. 86, pp. 360-371, 2018.
- S. Dabiri, C.-T. Lu, K. Heaslip, and C. K. Reddy, "Semi-supervised deep learning approach for transportation mode identification using GPS trajectory data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 5, pp. 1010-1023, 2019.
- 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.
- T. Bantis and J. Haworth, "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, pp. 286-309, 2017.
- T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785-794.
- V. N. Vapnik, "The nature of statistical learning," *Theory*, 1995
- W. Bohte and K. Maat, "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, pp. 285-297, 2009.
- W. S. Noble, "What is a support vector machine?" *Nature biotechnology*, vol. 24, no. 12, pp. 1565-1567, 2006.
- X. Song, H. Kanasugi, and R. Shibasaki, "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*, 2016, pp. 2618-2624.
- Y. Sasaki, "The truth of the F-measure," *Teach tutor mater*, vol. 1, no. 5, pp. 1-5, 2007.
- Y.-J. Byon and S. Liang, "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, pp. 264-272, 2014.
- Z. Huang, P. Wang, and Y. Liu, "Statistical characteristics and transportation mode identification of individual trajectories," *International Journal of Modern Physics B*, vol. 34, no. 10, p. 2050092, 2020.
- Z. Xiao, Y. Wang, K. Fu, and F. Wu, "Identifying different transportation modes from trajectory data using tree-based ensemble classifiers," *ISPRS International Journal of Geo-Information*, vol. 6, no. 2, p. 57, 2017.