


Article

Understanding Travel Mode Choice Behavior: Influencing Factors Analysis and Prediction with Machine Learning Method

Hui Zhang ^{1,*} , Li Zhang ¹, Yanjun Liu ¹ and Lele Zhang ²¹ School of Transportation Engineering, Shandong Jianzhu University, Jinan 250101, China² Yantai Yishang Electronic Technology Co., Ltd., Yantai 264003, China

* Correspondence: zhanghui@sdjzu.edu.cn

Abstract: Building a multimode transportation system could effectively reduce traffic congestion and improve travel quality. In many cities, use of public transport and green travel modes is encouraged in order to reduce the emission of greenhouse gas. With the development of the economy and society, travelers' behaviors become complex. Analyzing the travel mode choices of urban residents is conducive to constructing an effective multimode transportation system. In this paper, we propose a statistical analysis framework to study travelers' behavior with a large amount of survey data. Then, a stacking machine learning method considering travelers' behavior is introduced. The results show that electric bikes play a dominant role in Jinan city and age is an important factor impacting travel mode choice. Travelers' income could impact travel mode choice and rich people prefer to use private cars. Private cars and electric bikes are two main travel modes for commuting, accounting for 30% and 35%, respectively. Moreover, the proposed stacking method achieved 0.83 accuracy, outperforming the traditional multinomial logit (MNL) mode and nine other machine learning methods.

Keywords: travel mode choice; machine learning; travel behaviors; feature importance



Citation: Zhang, H.; Zhang, L.; Liu, Y.; Zhang, L. Understanding Travel Mode Choice Behavior: Influencing Factors Analysis and Prediction with Machine Learning Method. *Sustainability* **2023**, *15*, 11414. <https://doi.org/10.3390/su151411414>

Academic Editors: Sarbast Moslem, Tiziana Campisi, Nurten Akgün and Kh Md Nahiduzzaman

Received: 24 June 2023

Revised: 20 July 2023

Accepted: 21 July 2023

Published: 23 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Understanding the structure of travel modes among urban travelers plays a critical role in traffic planning, traffic management, traffic flow prediction, etc. The structure of travel modes is complicated and changeable due to many factors, such as population aging, economic growth and emerging transport modes [1–3]. In recent years, use of low-carbon travel modes, like public transport, bikes and walking, has been encouraged to reduce carbon emissions [4,5]. Previous studies have shown that there are many factors influencing the travel mode choices of travelers, like travel time, convenience, reliability, age, cost, the built environment and even epidemic diseases [6–9]. However, accurate prediction of the travel modes of urban residents according to sociodemographic and travel characteristics in terms of age, income, travel distance, etc., is a big challenge.

Generally, travelers' mode choice is affected by various factors, including traffic facilities, public transport service quality and personal attributes. In many areas, transit-oriented development is encouraged by the provision of better travel services [10,11]. With the help of advanced information technology, customized shuttle transit is developed rapidly, which could reduce waiting times and provide personalized service [12]. Mobility as a service (MaaS) is a hotspot that incorporates multiple traffic modes to improve travel quality. MaaS attracts individuals who engage with public transport services, while private car users do not significantly adopt it [13]. Detecting travel mode choices is conducive to promoting the development of MaaS. In many countries, population aging is becoming serious and the number of elderly people is increasing rapidly. Studies show that young people and the elderly have different travel behavior [14].

Mode choice is an important step in traditional four-step modeling. Discrete choice models (DCMs), such as the binary logit (BL) model [15], MNL model [16], nested logit

(NL) model [17] and mixed multinomial logit model [18], have been widely adopted to analyze travel mode choice behavior. In the majority of discrete choice models, individuals are typically assumed to be rational agents who select the option that offers the greatest utility. The utility is a function of both individual features and stochastic terms. Therefore, discrete choice models cannot offer exact predictions of the choices of individuals but assess the probability of each person choosing each substitute [19]. Structural equation modeling (SEM) can consider both observed and latent indicators, and it is a powerful tool to construct choice models [20–22]. Recently, the SEM-logit model has drawn much attention in the study of choice behavior by considering latent variable factors [23,24].

In recent years, machine learning methods based on big data have been widely applied in transportation fields, and they show high predictive performance that surpasses classic discrete choice models [25–27]. Existing studies have limitations. For example, the machine learning methods used are classic ones, which can outperform logit models. However, there is little research focused on mixing machine learning methods to enhance prediction accuracy. This paper tries to bridge these gaps. This paper compares the performances of the traditional MNL model and 10 machine learning methods: the support vector machine (SVM) method, classification and regression trees (CARTs), the random forest (RF) method, the gradient boosting decision tree (GBDT) method, adaptive boosting (AdaBoost), extreme gradient boosting (XGBoost), the light gradient boosting machine (LightGBM) method, hard voting, soft voting and stacking.

The contributions of this paper include three parts:

- It analyzes the characteristics of travelers and their behaviors;
- It provides a comprehensive introduction to machine learning classifiers;
- It proposes the stacking model, with which predictive performance is the best.

The rest of the paper is organized as follows. Section 2 introduces the relevant literatures. Section 3 introduces the methodology. Section 4 provides the data description. Section 5 provides the characteristics of the travel modes. Section 6 describes the model evaluation and comparison. Finally, Section 7 concludes the paper.

2. Literature Review

Travel mobility is a key component of travel characteristics that is conducive to understanding travelers' behaviors. Previous studies mainly refer to model-based and data-driven methods. Model-based methods are concentrated on the structural equation model, discrete choice model and latent class model. Rafiq and McNally [28] used SEM to investigate the work tour behavior of transit commuters and determined that married men in low-density areas who own vehicles and have no children have a preference for simple work tours. Ding et al. [6] proposed an integrated structural equation model and DCM and found that higher accessibility is correlated with higher likelihood of choosing walking or cycling and distance to transit has a crucial effect on choosing public transport. Molin et al. [29] studied the multimodal travel groups and attitudes of Dutch travelers using a latent class cluster method and found that solo car drivers have more negative attitudes to transit and bikes. Zhang et al. [30] proposed a two-level nested logit model to examine the travel behavior of elderly individuals in Beijing and concluded that buses play a very significant role in motorized travel for this demographic.

Recently, data-driven methods have drawn much attention in detecting travel behaviors. Automatically collected big data (e.g., taxi GPS data, smart card data, bike-sharing data and mobile data) provide new insights into travel mobility and have been widely applied in transportation systems [31–34]. Faroqi et al. [35] measured the activity similarity among public transit riders based on smart card data and found that activity patterns are different on weekends than on weekdays. Zhang et al. [36] used taxi trajectory data to explore travel demand by grid divisions. Mobile phone data could record consecutive trajectories, which is valuable data for the study of travel behaviors. Bachir et al. [37] proposed a two-step semi-supervised learning method to identify transport modes and then studied the dynamic origin–destination flows.

There are various factors that influence travelers' mode choices, like age, cost, travel time, etc. Studies show that travelers in households with more children are more likely to choose transit, walking or cycling [6]. Truong and Somenahalli [38] claimed that elderly people who avoid driving at peak hours are more likely to utilize public transport. Age was found to play a negative role in public transport usage for elderly people in most Western countries [39]. Jin and Yu [40] found that women 40 years old or more were less likely to use transit compared to men. Sun et al. [41] pointed out that the proportion of four-way intersections, road density and population density in residential areas are negatively correlated with driving probability. Nguyen et al. [42] claimed that senior people, students and unskilled laborers prefer to use buses. Mao et al. [43] found that there is a significant portion of commuters who show multimodal behaviors.

Logit-based models are the most popular model to study travelers' travel mode choices. Ma et al. [44] proposed a nested logit joint model and found that rich people have a greater probability of commuting by taxi or driving alone. Li et al. [45] comparatively analyzed travel mode choice using logit and weibit models and showed that males prefer to travel by car more than females. Al-Salih and Esztergar-Kiss [46] proposed a multinomial logit model and a nested logit model with a utility function and found that trip distance was the most significant variable. The structural equation model is another widely used model to analyze relationships between variables. Faber et al. [47] found that public transport use is most strongly influenced by the built environment, while car and bike use is hardly influenced. Shin and Tilahun [48] studied travel behaviors of young adults using SEM and found that young households were more likely to live closer to the core of cities. The emerging machine learning methods have been proven to have better performance in analyzing travel mode choices. Zhao et al. [27] found that the random forest model exhibits notably higher prediction accuracy compared to the multinomial logit and mixed logit models.

In recent years, machine learning methods have attracted much attention for analyzing travel mode choice. Pineda-Jaramillo and Arbelaez-Arenas [49] compared different logit models and machine learning models and found that the optimized gradient boosting model outperformed the compared models. Mi et al. (2021) proposed a softmax regression machine learning algorithm to predict travel mode and the results showed that the prediction accuracy was higher than with the SVM and multinomial logit (MNL) models [50]. Salas et al. (2022) provided a systematic comparative evaluation of machine learning and discrete choice models and found that neural networks outperform other models, including the MNL, mixed multinomial logit (MMNL), K-nearest neighbor (KNN), SVM, random forest (RF) and XGBoost models [19]. Kashifi et al. (2022) proposed five machine learning methods to understand travelers' mode choices and found that the light gradient boosting decision tree (LightGBDT) model surpassed other models [51]. Xia et al. (2023) developed a random effect Bayesian neural network (RE-BNN) framework considering the regional heterogeneity of travel behaviors to model travel mode preferences [52]. Elharoun et al. (2023) proposed deep neural networks (DNNs) to analyze the travel mode choices of people in Mansoura, Egypt, and found that they outperformed other machine learning classifiers [53]. Chen and Cheng (2023) proposed combinations of statistical/machine learning methods and six over/undersampling techniques to improve prediction performance [54].

The former studies show that the prediction accuracy of the MNL models is not high. These discrete choice models could not give exact predictions of which mode people would choose, but they gave the probability of each person's choice. The recent studies show that machine learning is an effective alternative to predict travel modes. However, finding a more effective method is also challenging because the travel mode choice is complex. To fill this gap, this paper uses a stacking ensemble machine learning method, which combines different methods to create a powerful framework to enhance prediction accuracy.

3. Methodology

In this part, we introduce models used to predict different modes based on the characteristics of travelers' trips. The methods of the multinomial logit model and seven single learning models are described. Then, ensemble machine learning classifiers are introduced.

3.1. Discrete Choice Model—Multinomial Logit (MNL) Model

The multinomial model is one of the most widely used discrete choice models for travel mode predictions. The MNL model assumes that the random components of the utility functions are independent of each other and obey the same extreme-value Gumbel distribution. Therefore, we suppose that individuals make their travel mode choices based on rational considerations; i.e., that they select a trip mode that will give them the highest utility. The utility function is shown in Equation (1).

$$T_{ij} = x_i\beta_j + \varepsilon_{ij}(i = 1, 2, \dots, n), \quad (1)$$

where T_{ij} is the utility of the choice of mode j by the individual i , x_i is the vector of individual attributes (e.g., sex, profession, trip distance), β_j is the vector of the parameters to be estimated and ε_{ij} is the unobserved random error.

In terms of utility maximization and random utility theory, the individual i will choose the mode j if $T_{ij} > T_{im}$. The probability of individual i choosing mode j is given by Equation (2).

$$P_i(j) = P(T_{ij} > T_{im}), (m \neq j), \quad (2)$$

where m represents another mode.

Assuming that the error term ε_{ij} is Gumbel-distributed and substituting Equation (1) into Equation (2), the probability of an individual i choosing the alternative j is obtained as shown in Equation (3).

$$P_i(j) = \frac{\exp(x_i\beta_j)}{\sum_{m=1}^J \exp(x_i\beta_m)}, \quad (3)$$

where J is the alternatives set.

3.2. Single Machine Learning Classifiers

In contrast with logit models, many machine learning models are non-parametric, which allows for a more flexible structure from the training data.

(1) Support vector machine

The SVM model is a linear binary classifier that finds a hyperplane capable of dividing the set of dissimilar data to the maximum extent. Its segmentation strategy is interval maximization, which can be transformed into an optimization algorithm for convex quadratic programming. Also, SVM models using a kernel function can solve nonlinear separable problems. The kernel function maps the problem from the original data to higher dimensions [55], where it is feasible to linearly separate training data into multiple groups. There are multiple hyperplanes that can separate the data, and the separating hyperplane with the largest geometric interval for the training set is called the maximum margin. Such a hyperplane that generalizes well has a better classification prediction ability for data because it not only classifies the data correctly but also separates the points that are close to the hyperplane with high enough certainty. It is possible to achieve multiclass classification by training a large number of binary classifiers, such as the one-vs-rest, one-vs-one, oriented acyclic graph and binary tree approaches.

(2) Decision tree

The decision tree is a kind of decision from a root node to a leaf node consisting of nodes and directed edges. The types of nodes are divided into two: inner nodes, which represent a test condition for a feature or attribute, and leaf nodes, which represent a classification. Different decision trees are generated according to different selection

branching functions, such as ID3, C4.5 and CARTs. CARTs work by dividing the attributes based on the Gini index, recursively dividing the sample into two subsets and finally generating a binary decision tree. The split criteria of ID3 and C4.5 are respectively the information gain and information gain ratio and involve a large number of logarithmic operations. The CART algorithm, which uses another metric called the Gini index to split the nodes, simplifies the model and retains the advantages of the entropy model. Equation (4) describes the Gini coefficient.

$$Gini(t) = 1 - \sum_{n=1}^N [P(n|t)]^2, \quad (4)$$

where $P(n|t)$ is the probability of the node t being classified to the class n , and N is the number of classes.

3.3. Ensemble Learning

In ensemble learning, the results of a variety of different classifiers are merged to take advantage of their complementarity with the hope that the entire ensemble prediction will be more accurate than the individual ones [56].

This algorithm can be broadly broken into two steps. The first is to build a range of individual learners. The second is to define a combination strategy to integrate these individual learners. Individual learners can generally be arbitrary machine learning models; e.g., decision tree models and support vector machine models.

There are two categories, homogeneous ensembles and heterogeneous ensembles, depending on the individual learners. Members of the homogeneous ensemble all use a single type of base learning method [57]. Bagging and boosting are two popular homogeneous ensemble models. On the other hand, members of the heterogeneous ensemble use different base learning algorithms. Voting and stacking are the commonly used methods in homogeneous ensembles.

(1) Bagging

Bagging, also known as bootstrap aggregation, is a parallel ensemble learning technique that is commonly used to reduce error by implementing a set of homogeneous ensemble methods. The core concept of bagging is to train multiple base learners separately with resampled training data to create a more accurate and stable model.

The random forest method is a bagging method that uses both bagging and feature randomness to construct a multitude of decision trees [58]. The CART model is the base learner of the random forest method. At first, N training samples (the training set size is N) are selected randomly and retained as the training set of each tree. This sampling method is called bootstrap sampling. Then, a subset of m ($m < M$) features (the feature dimension is M) is selected. Each tree grows to the maximum extent possible and there is no pruning process. The prediction result for classification is determined by majority voting.

(2) Boosting

Boosting is a serial ensemble learning algorithm that combines multiple weak learners in a sequence to form a strong learner. The key idea of boosting is to reassign the weights of samples, which gives more attention to the observations wrongly classified in the previous weak learner, to train the next weak learner [59]. Some popular types of boosting methods include AdaBoost, GBDT, XGBoost and LightGBM.

AdaBoost was the first boosting algorithm that could adaptively tweak the subsequent modeling process based on the results of the previous weak learner and consider the output of all weak learners. First, a decision tree is built from the full training sets. Then, the training weights are adaptively modified to give instances that mistakenly predict more weight and instances that properly predict less weight. The process goes on until the specified condition is met. The prediction result is the weighted average value of all weak learners, whose weight is also computed adaptively in the above process.

GBDT is similar to AdaBoost, but it influences the subsequent weak learners by fitting the residuals instead of modifying the sample weights [60]. More types of loss functions can be used by GBDT. XGBoost and LightGBM are wide implementations of these methods with significant extensions to enhance prediction accuracy and computing speed [61].

GBDT fits the residuals by using the negative gradient of the model on the data as an approximation of the residuals. XGBoost also fits the residuals on the data, but it approximates the model loss residuals with a Taylor expansion, while XGBoost improves the loss function of the model and adds a regular term for the model complexity.

Like XGBoost, LightGBM has many advantages, including sparse optimization, parallel training, multiple loss functions, regularization, bagging and early stopping. The major difference between the two methods is in the order in which the tree is expanded. XGBoost grows trees level-wise and row by row, as most other implementations do. In contrast, LightGBM grows trees leaf-wise (best first). It chooses the leaf with the maximum delta loss to grow.

(3) Voting

There are two kinds of voting models: hard voting and soft voting. Hard voting is defined as counting the number of occurrences of different categories, and the category with the highest number of occurrences is the voting result. Soft voting involves summing the probabilities under different categories, and the category with the highest probability is the voting result.

(4) Stacking

As an emerging algorithm, the stacking ensemble learning method has been applied in different fields with good results. The special feature of stacking is that it can integrate different kinds of models. For solving the classification problem, we present a stacking integrated framework with two layers as an example. In general, the first layer of the framework consists of multiple base classifiers, and the input of each base classifier is the training data. The second layer of the framework is the meta-classifier. The training data of the meta-classifier consist of the output of the first base classifier and the label of the original training data. Then, the meta-classifier is trained to get the final results. The structure of the stacking ensemble learning method is shown in Figure 1.

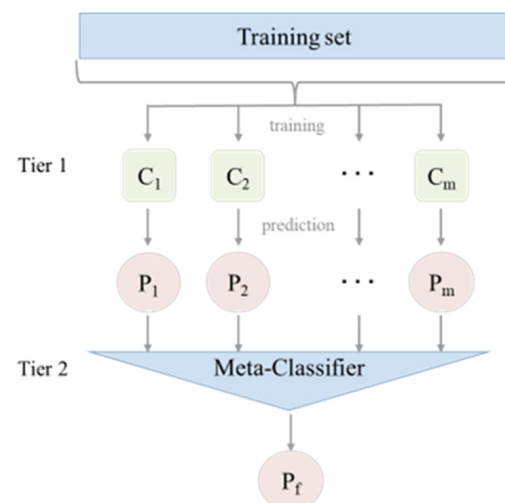


Figure 1. Stacking ensemble learning method.

4. Data Description

The data used in this study were the 2019 household travel survey data for Jinan, China, which represent 19,551 households, 38,133 members and 80,101 trips. Jinan is the capital city of Shandong province, which lies in the east of China. The household survey

data comprise four kinds of information: household information, household members' information, information for trips made within the past 24 h for each household member and transit-related information for each member. The spatial distribution of the trips of the survey travelers is shown in Figure 2.

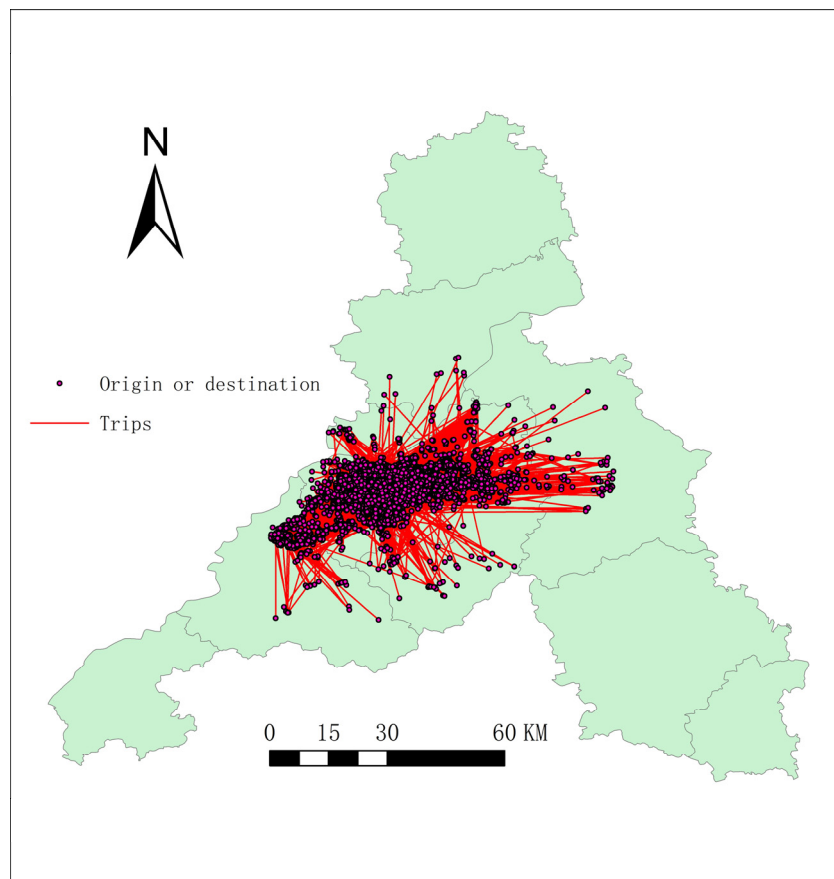


Figure 2. Spatial distribution of surveyed travelers and OD trips.

In this study, we merged the four kinds of information into one table, and the main fields are exhibited in Table 1. The travel modes related to the survey were bus, private car (PC), walking, taxi, electric bike (EB), motorcycle (MC), bike sharing (BS), subway and others. The travel purposes were investigated, which included commuting, going home, entertainment, shopping and dinner (SD), official business (OB), school, tour, picking up children (PUC), visiting friends (VF) and others.

Table 1. Main information for survey data.

Fields	Example
Traveler ID	1
Family ID	2
Gender	Male
Age	32
Administrative district	Lixia
Number of family members	2
Number of temporary residents	0
Home address	Xindongfang garden
Longitude of home address	117.118248

Table 1. *Cont.*

Fields	Example
Latitude of home address	36.691411
Travel costs of family	CNY 500–899
Number of fixed parking spaces	1
Number of cars	1
Number of motorcycles	0
Number of bicycles	0
Number of electric bikes	0
Occupation	Teacher
Work address	Xuyang Corporation
Longitude of work address	117.050012
Latitude of work address	36.661655
Person's monthly income	CNY 3000–5000
Departure time	8:00
Departure address	Yajuyuan
Longitude of departure address	117.125085
Latitude of departure address	36.691915
Arrival time	8:20
Arrival address	Longao building
Longitude of arrival address	117.130894
Latitude of arrival address	36.676144
Trip purpose	Working
Trip distance	2.56 km
Travel mode	Bus
Travel time	20 min
Frequency of transit use per week	5
Monthly transportation expenditure of family	CNY 500
Walking time to transit stop	5 min
Census register	Local household registration

The data were preprocessed before use. The preprocessing was as follows: (1) the trips with missing information were removed; (2) the trips that were outside of the studied areas were removed; and (3) duplicated and erroneous data were removed.

5. Multimodal Travel Characteristics

In this study, we extracted the main attributes of travelers, including age, gender, personal monthly income (PMI), occupation, monthly transportation expenditure of the family (MTEF), frequency of transit use (FTU), number of family members (NFM), number of cars (NC), number of parking places (NPP), travel distance, travel mode, travel purpose, travel time and walking time to transit stop (WTTS). The data consisted of both continuous variables and a significant number of categorical variables. Furthermore, most of these categorical variables exhibited meaningful ordinal relationships among their categories. Tree models, which are commonly employed in the later stages of machine learning methods, typically do not require one-hot encoding for categorical variables, so this study adopted a label encoding approach for handling the categorical variables. In addition, for some variables, data binning was used to further enhance the representation of information within the dataset. We quantified some attributes that were not discrete values originally, as shown in Table 2. For example, we used “1” to represent males and “2” to represent females.

5.1. Travel Distances for Different Travel Modes

Figure 3 shows a violin plot of travel distances for different travel modes. The numbers in the figure are the total numbers of travelers using different travel modes. In Figure 3, buses, private cars and others show large ranges, and they are concentrated between 0 and 10 km. The average values for the travel distance were 5.19, 6.58 and 10.75 for buses, private cars and others. Moreover, the walking travel mode exhibited a small range and the average travel distance was 0.58 km. In addition, electric bikes were the most popular travel mode with the largest number of travelers. The average travel distance of an electric bike is 2.49 km, which is larger than for bike sharing and walking. In most cities in China, electric bikes have become a crucial travel mode due to their convenience and flexibility. Bike sharing is an important travel mode to connect transit stations. Most bike-sharing trips' distances are within 5 km.

Table 2. Description of basic indicators.

Variables	Numeric Value
Gender	Male: 1; female: 2
Age	(0, 18]: 1; (18, 24]: 2; (24, 40]: 3; (40, 60]: 4; (60, 70]: 5; (70, ∞]: 6
Personal monthly income (PMI)	(CNY 0, 3000]: 1; (CNY 3000, 5000]: 2; (CNY 5000, 7000]: 3; (CNY 7000, 10,000]: 4; (CNY 10,000, 15,000]: 5; (CNY 15,000, ∞): 6
Occupation	Clerk: 1; service industry employees: 2; science, education, culture and health workers: 3; government, enterprises and institutions: 4; workers: 5; retired people: 6; self-employed: 7; students: 8; unemployed people: 9; military and police: 10; production personnel in agriculture, forestry, animal husbandry, fishery and water conservancy: 11; other: 12
Monthly transportation expenditure of family (MTEF)	(CNY 0, 100]: 1; (CNY 100, 500]: 2; (CNY 500~900]: 3; (CNY 900, 1400]: 4; (CNY 1400~1800]: 5; (CNY 1800~2200]: 6; (CNY 2200, 2600]: 7; (CNY 2600, ∞): 8
Travel distance	(0, 5 KM]: 1; (5 KM, 10 KM]: 2; (10 KM, 15 KM]: 3; (15 KM, 20 KM): 4; (20 KM, 30 KM]: 5; (30 KM, ∞): 6
Travel mode	Bus: 1; private car: 2; walking: 3; taxi: 4; electric bike: 5; motorcycle: 6; bike sharing: 7; subway: 8; others: 9
Travel purpose	Commuting: 1; going home: 2; entertainment: 3; shopping and dinner: 4; official business: 5; school: 6; tour: 7; picking up children: 8; visiting friends: 9; others: 10
Travel time	(0, 10 min]: 1; (10 min, 20 min]: 2; (20 min, 30 min]: 3; (30 min, 45 min]: 4; (45 min, 60 min]: 5; (60 min, ∞): 6
Walking time to transit stop (WTTS)	(0, 5 min]: 1; (5 min, 10 min]: 2; (10 min, 15 min]: 3; (15 min, 20 min]: 4; (20 min, ∞): 5
Frequency of transit use per week (FTUPW)	Actual value
Main travel mode during the weekday (MTMDW)	Actual value
Number of family members (NFM)	Actual value
Number of temporary residents (NTR)	Actual value
Administrative district	Actual value
Census register	Actual value
Number of fixed parking spaces (NFPS)	Actual value
Number of cars (NC)	Actual value
Number of bicycles (NB)	Actual value
Number of electric bikes (NEB)	Actual value
Number of motorcycles (NM)	Actual value

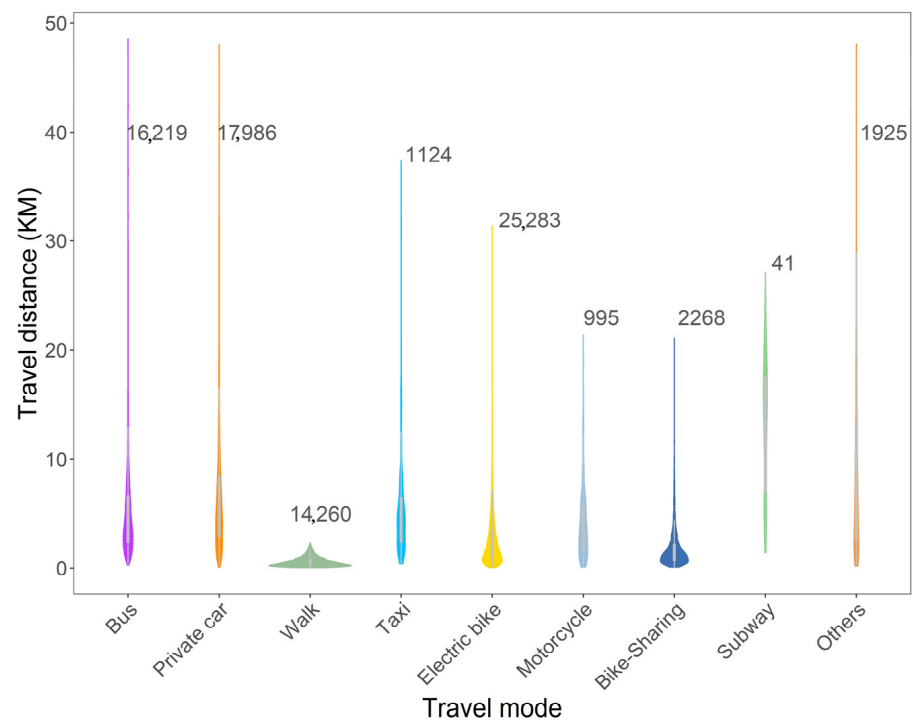


Figure 3. Violin plot of travel distances for different travel modes.

5.2. Travel Flow Distributions of Different Travel Modes

To explore the temporal characteristics, we plotted the travel flow distributions of different travel modes in a day, as shown in Figure 4. The unit of the vertical axis is the number of travelers. We can see that most kinds of travel modes have two peaks in the morning (7:00–9:00) and afternoon (17:00–19:00). During the off-peak hours, trips by bus, electric bike and private car decrease sharply. However, we found that taxis have no obvious travel peaks compared with other travel modes because taxi users seldom travel for commuting.

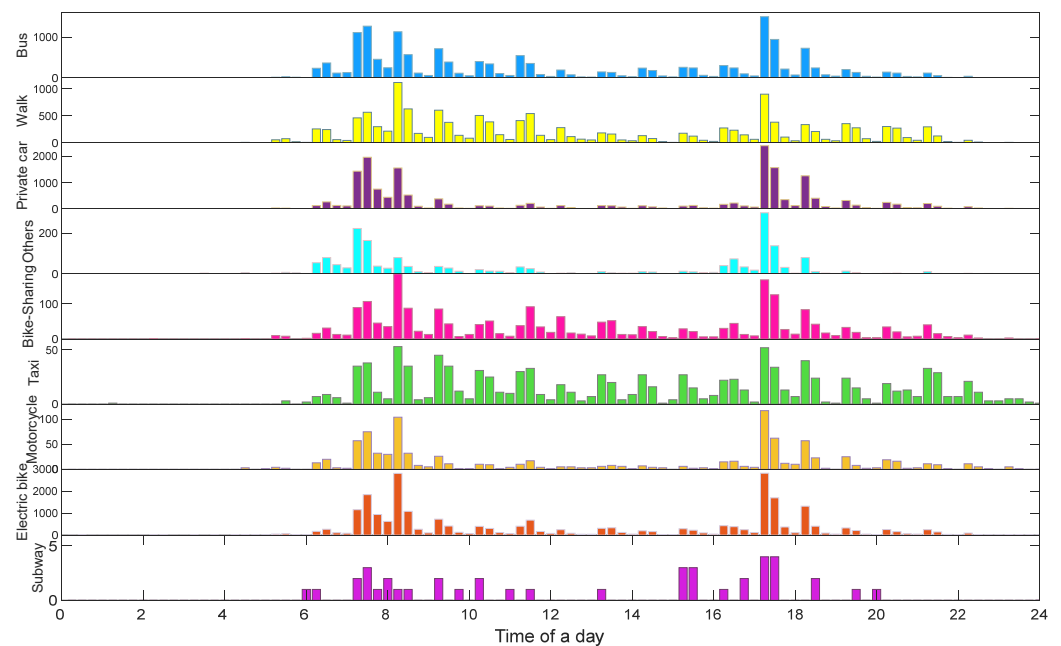


Figure 4. Travel flow distributions of different travel modes in a day.

5.3. Distributions of Basic Features

To understand the basic statistical characteristics of urban travelers, we plotted the distributions of indicators, as shown in Figure 5. Travelers' age played a key role in traffic mode choice and other travel behaviors. Studies have shown the travel behaviors of the elderly (60+ years old) and younger people [62]. As can be seen in Figure 5b, 16.34% of people were elderly people, which implies that the city has an aging society. People's income could impact the travel mode choice, and rich people preferred to use private cars. Figure 5c shows that most people's income was concentrated between CNY 0 and 7000, and a small proportion of people earned more than CNY 1000. Figure 5e shows the distribution of monthly transportation expenditure by families, which exhibited a lognormal distribution. Figure 5h shows the distribution of travel purposes, indicating that "commuting" and "go home" were the main travel purpose, which accounted for 30.57% and 46.52%. Public transport is a green travel mode with low costs, which encourages many people to use it. Figure 5i indicates that most people could walk to the transit stop within 10 min. Figure 5j,k show that most families owned private cars and parking places, respectively.

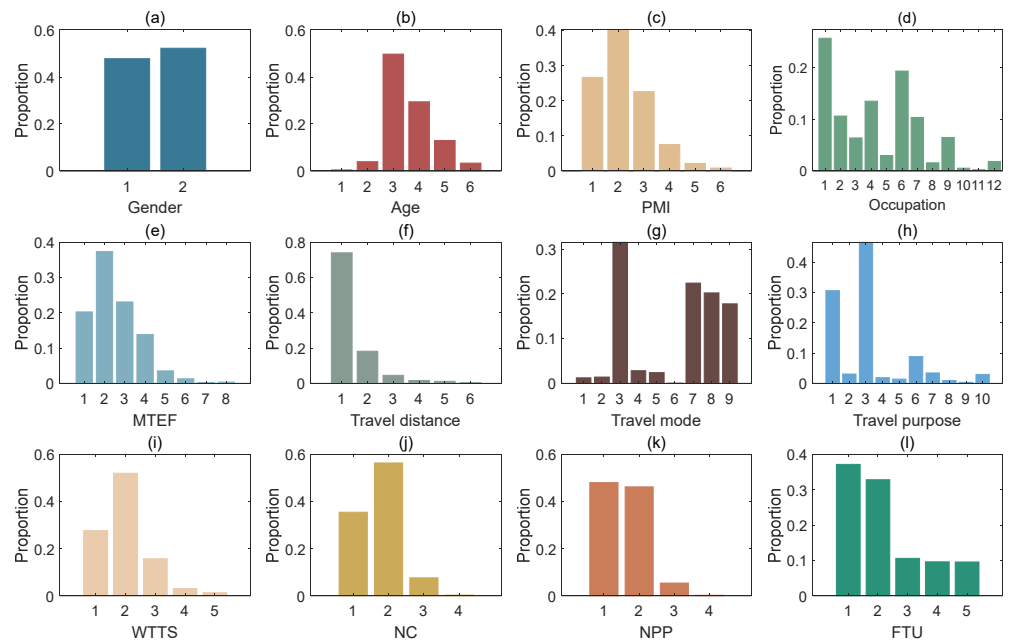


Figure 5. Distributions of main indicators.

5.4. Travel Purposes Associated with Different Travel Modes

Travel purpose played a significant role in the travel mobility analysis. Table 3 shows the statistical results for the trip distance and proportions of travel modes with different travel purposes. Commuting was an essential travel purpose. Private cars and electric bikes were two main travel modes for commuting, accounting for 30% and 35%, respectively. Walking played a dominant role in entertainment and SD, while private cars were the primary choice for official business. Bus was the main option for school and VF. We can see that travelers had different predilections for different travel purposes.

Table 3. Trip distance and travel mode distributions for different purposes.

Trip Purpose	Trip Distance			Proportions of Travel Modes								
	Min.	Max	Median	Bus	PC	Subway	Walk	BS	EB	MC	Taxi	Others
Commuting	0.05	48.5	3.24	0.17	0.30	0.002	0.086	0.03	0.35	0.02	0.002	0.04
Going home	0.05	48.5	2.44	0.20	0.23	0.004	0.17	0.03	0.32	0.01	0.006	0.03
Entertainment	0.05	39.8	1.06	0.21	0.04	0.005	0.51	0.03	0.17	0.01	0.005	0.02
SD	0.05	45.9	1.09	0.21	0.09	0.003	0.34	0.05	0.27	0.01	0.007	0.02

Table 3. Cont.

Trip Purpose	Trip Distance			Proportions of Travel Modes								
	Min.	Max	Median	Bus	PC	Subway	Walk	BS	EB	MC	Taxi	Others
OB	0.06	45.4	3.71	0.15	0.37	0.005	0.07	0.03	0.19	0.02	0.035	0.13
School	0.15	34.2	3.97	0.56	0.11	0.004	0.09	0.04	0.13	0.01	0.016	0.04
Tour	0.07	46.5	2.02	0.38	0.05	0.001	0.35	0.03	0.16	0.01	0.009	0.01
PUC	0.06	48.0	1.11	0.10	0.19	0.000	0.178	0.01	0.5	0.01	0.002	0.01
VF	0.06	42.4	3.23	0.42	0.13	0.001	0.10	0.03	0.26	0.01	0.030	0.01
Others	0.05	46.1	1.98	0.32	0.14	0.002	0.22	0.03	0.22	0.01	0.018	0.04

5.5. Correlations of All Variables

The Pearson correlation coefficient, computed with Equation (5), was employed to explain in detail the correlations between the variables. The results are shown in Figure 6, where darker colors represent a closer relationship and lighter colors represent the opposite.

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \tag{5}$$

where r_{xy} is the Pearson correlation coefficient, n is the sample size, x_i and y_i are the i -th sample point and \bar{x} and \bar{y} are the sample mean.

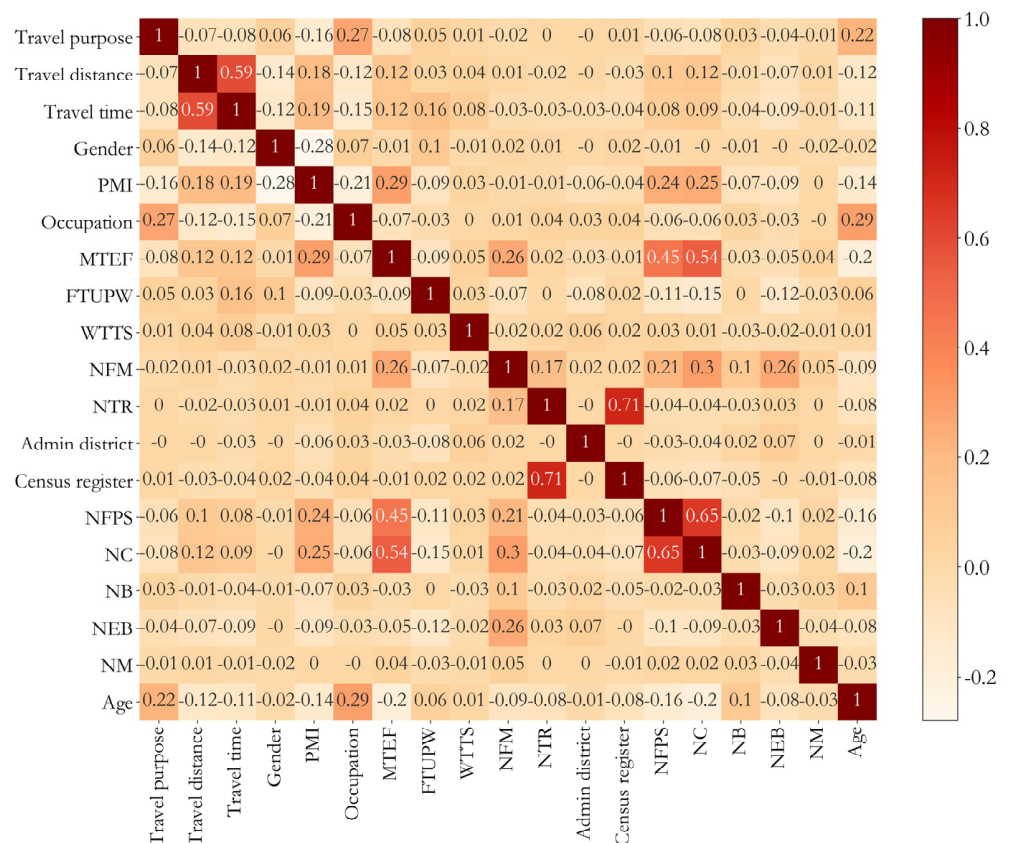


Figure 6. The correlations of all variables.

We can see that the census register was more highly correlated with the number of temporary residents. Numbers of temporary residents in the family were greater for local households. The relationship between the number of cars and the number of fixed parking spaces was the same, and they were complementary to each other.

6. Model Evaluation and Comparison

In this section, we propose a range of models to predict the travel modes of travelers based on their attributes, including MNL, SVM, Adaboost, CART, GBDT, random forest, LightGBM, XGBoost, hard voting, soft voting and stacking models. This section presents the model development. Specifically, we evaluated the model performance and compared different models.

6.1. Model Development

At first, the p -values of each feature were calculated. All values were lower than 0.05, which proved that they were closely related to the labels. To avoid high-intensity correlations among the variables, we also calculated the variance inflation factor (VIF) and found that all values were less than 10. Therefore, no variables needed to be removed. The logit model was developed using SPSS 26.0, and other machine learning models were produced with Python 3.8.5.

For machine learning methods, the dataset was split into a training subset (70% of the data) and a testing subset (30% of the data). Specifically, the training set consisted of 56,070 samples, while the testing set comprised 24,031 samples. To correct the imbalance in the data, the proportions of values in the training set and testing set were the same as the proportion of the samples provided. All prediction models used the same training set and test set. The training dataset was used to fit models and the test dataset provided a benchmark for assessing the models. The scikit-learn package's GridSearchCV module was used to tune and produce the best combination of hyperparameters in each model based on a scoring metric.

To reduce variability and avoid overfitting, we used fivefold cross-validation to evaluate the skills of these models. Specifically, the training sample was randomly partitioned into five equal-sized subsets. In each round, one subsample was kept as the validation dataset for testing the model, and the remaining four subsamples were used as a training dataset. After five iterations, each observation had been used precisely once for validation and each observation had been utilized for training. The last result comprised the average values for each process.

For the ensemble machine learning methods, the SVM model and XGBoost model were used to combine voting methods. For stacking models, the specific steps are shown in Figure 7. The stacking model has a multi-layer structure, and the output of each layer is used as input to the next layer. Therefore, the more layers are constructed, the more complex the model is and the slower the training is. In this study, the stacking model we used was divided into two layers. The first layer consisted of the seven models introduced in Section 1; i.e., the AdaBoost, CART, GBDT, random forest, LightGBM, XGBoost, and SVM models. Each of these models was trained using specific parameters, as detailed in Table 4. After five-fold cross-validation, the output of each learner was gathered into new training sets and prediction sets. When constructing the new training and prediction sets, equal weights were assigned to each learner. Moving to the second layer, we employed the SVM model as a meta-learner to combine the new training and test sets generated by the first layer model for training and prediction and output the final prediction results.

Table 4. A binary confusion matrix.

		Predicted Class	
		Class = 1	Class = 0
Actual Class	Class = 1	True-positive (TP)	False-negative (FN)
	Class = 0	False-positive (FP)	True-negative (TN)

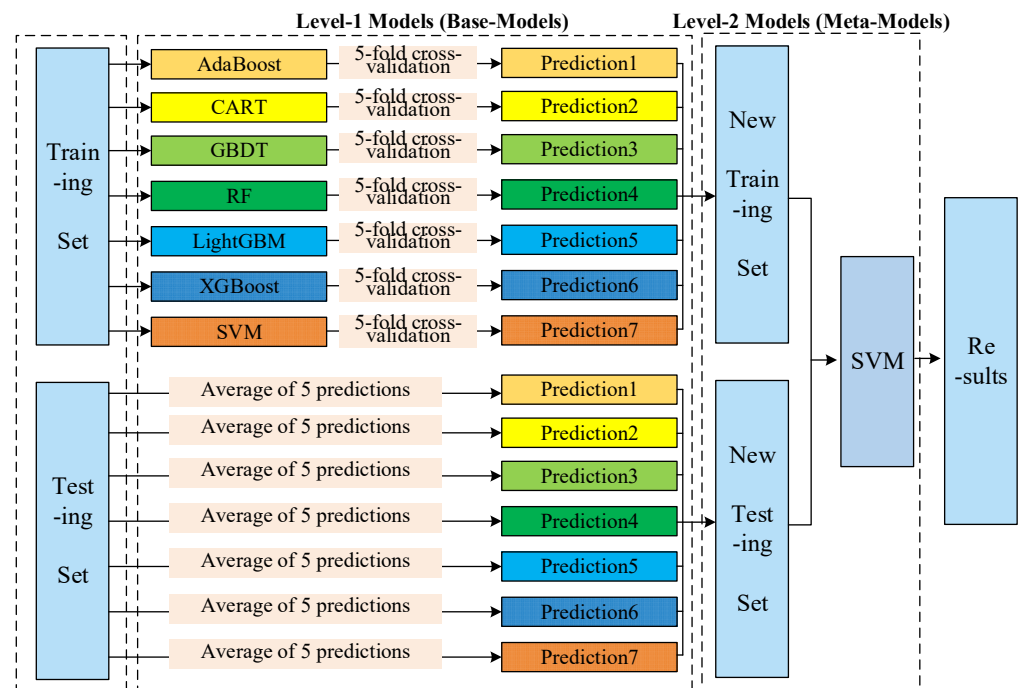


Figure 7. Stacking model in this study.

6.2. Predictive Accuracy Comparison

Various evaluation metrics were employed to comprehensively assess the performances of the multi-label classification models. Accuracy, precision, recall, F1-score, the Matthews correlation coefficient (MCC) and Cohen’s kappa coefficient (kappa) were utilized for this purpose in this study, and their corresponding equations are presented as Equations (6)–(12). The confusion matrix is a better way to evaluate predictive models when working with imbalanced datasets. Table 4 presents a binary confusion matrix along with the corresponding definitions of true-positive (TP), false-positive (FP), false-negative (FN) and true-negative (TN) classifications. It is a tabular summary that counts correct and incorrect predictions for each mode. In the confusion matrix, the correctly classified samples (actual label = predicted label) are distributed on the diagonal line from the top left to the bottom right. From the confusion matrix, the prediction accuracy can be defined as the ratio of the number of correctly classified (on the diagonal) samples to the total number of samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (10)$$

$$Kappa = \frac{(Accuracy - E)}{(1 - E)} \quad (11)$$

$$E = \frac{(TP + FP) \times (TP + FN) + (TN + FP) \times (TN + FN)}{(TP + TN + FP + FN)^2} \quad (12)$$

Table 5 shows the hyperparameters and performance criteria for each model. As shown by the comparison of the prediction performance of the models (see Figure 8 and Table 5), the stacking machine learning classifier performed the best among these models, with high accuracy, precision, recall, F1-scores, MCC and kappa. First, we can see that the logit model had a relatively low accuracy of 0.66. Accuracy represents the proportion of samples correctly classified by the classifier, indicating that this model could only correctly classify about 66% of the samples. Its precision was 0.58, recall was 0.62 and F1-score was 0.54, indicating mediocre performance. The Matthews correlation coefficient (MCC) and kappa coefficient were both close to 0.5, indicating that the model's performance was comparable to random classification. The SVM model had an accuracy of 0.79, significantly higher than the logit model. It had a precision of 0.74, recall of 0.62 and F1-score of 0.66, indicating relatively good performance. The MCC and kappa coefficient were 0.74, indicating that the model performed well. The Adaboost, CART, GBDT, and random forest models had accuracies ranging from 0.66 to 0.76. The Adaboost model had low precision, recall and F1-scores, with values of only 0.3, 0.31 and 0.29, indicating poor performance. The CART, GBDT and random forest models had relatively higher precision, recall and F1-scores, but there was still room for improvement. The LightGBM and XGBoost models had accuracies ranging from 0.76 to 0.78. LightGBM had high precision, recall and F1-scores, while the XGBoost model had a high precision of 0.8 but a lower recall. In terms of ensemble methods, the accuracies of the hard voting, soft voting and stacking models ranged from 0.79 to 0.86. Both the hard voting and soft voting models performed relatively well, but the stacking model performed the best. It had high precision, recall and F1-scores, and the MCC and kappa coefficient were close to 0.8, indicating excellent performance.

Table 5. Prediction models.

Models	Hyperparameters	Accuracy	Precision	Recall	F1-Score	MCC	Kappa
Logit model	MNL, C = 1	0.66	0.58	0.62	0.54	0.52	0.50
SVM	Decision function shape = "ovo", C = 10, gamma = 0.2, kernel = "rbf"	0.79	0.74	0.62	0.66	0.74	0.74
Adaboost	n estimators = 46, learning rate = 0.1	0.66	0.3	0.31	0.29	0.58	0.56
CART	Criterion = "gini", max. depth = 10	0.76	0.68	0.56	0.59	0.67	0.67
GBDT	n estimators = 46, min. sample leaves = 10	0.74	0.65	0.53	0.56	0.67	0.67
Random forest	n estimators = 16, min., samples split = 80, max. depth = 13, learning rate = 0.1	0.76	0.66	0.53	0.55	0.68	0.67
LightGBM	Feature fraction = 0.5, learning rate = 0.1, Max. depth = -1, num. leaves = 64	0.78	0.7	0.59	0.61	0.71	0.70
8 XGBoost	learning_rate = 0.1, gamma = 0, n_estimators = 28, max_depth = 10,	0.77	0.8	0.57	0.61	0.70	0.70
Hard voting	/	0.79	0.71	0.55	0.6	0.73	0.73
1 Soft voting	/	0.80	0.79	0.59	0.61	0.74	0.74
1 Stacking	/	0.86	0.83	0.62	0.67	0.78	0.78

The predictive results for each travel mode were different in these models. Figure 9 shows the performance of each travel mode in the stacking models. The model showed good performance in predicting private cars, electric bikes and motorcycles, with F1-scores of 0.9, 0.87 and 0.85, respectively. This indicates that the model's predictions were relatively accurate for these modes of transportation. In the case of predicting buses, the model also performed well, with an F1-score of 0.84. However, the model's performance was poor in predicting taxis, bike sharing and subways, with F1-scores of 0.45, 0.43 and 0.27,

respectively. The poor performance of the model in predicting taxis, bike sharing and subways can be attributed to the limited amount of data available for these categories and the imbalance in the samples.

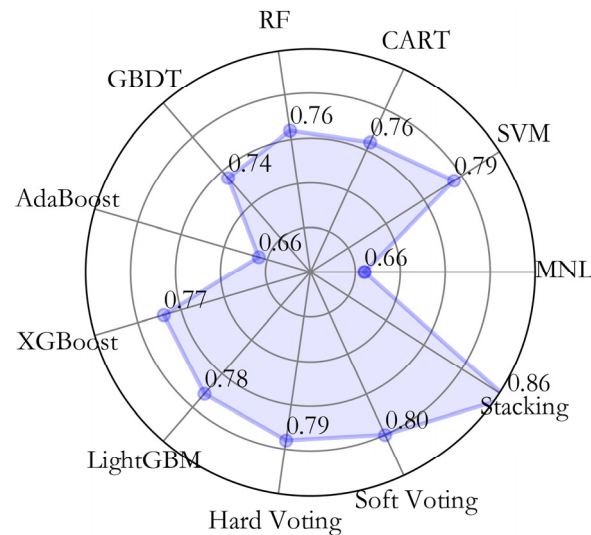


Figure 8. Comparison of accuracy of models.

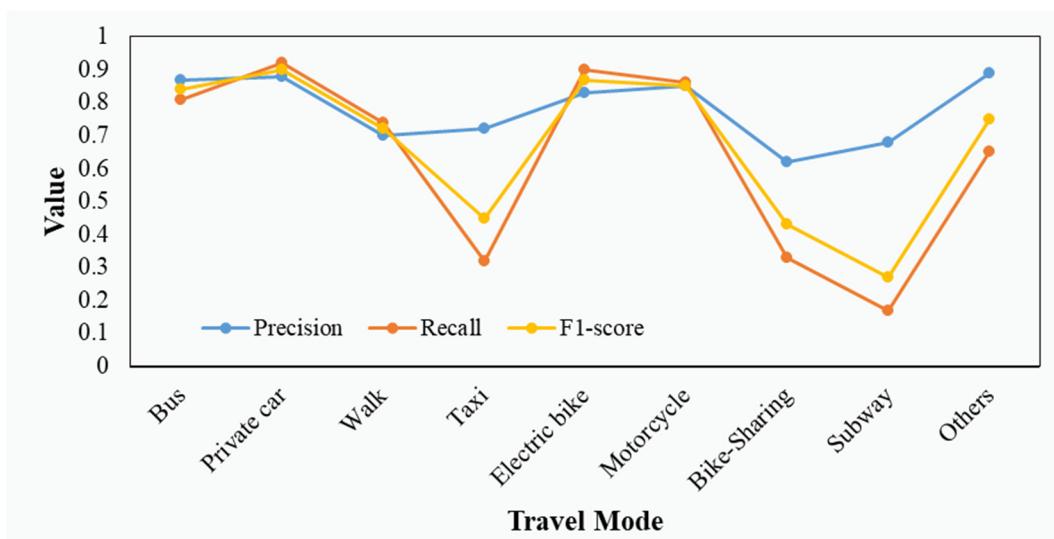


Figure 9. Stacking model performance for each travel mode.

6.3. Model Interpretation (Feature Importance)

Feature importance is a technique for assigning weights to input features based on their usefulness in predicting the target variable. In other words, it tells us which features can best predict the target variable. Models are explained well by feature importance.

The methods based on decision tree models can output feature importance while performing classification. The Gini index (see Equation (4)) was used to evaluate the importance of each feature in determining the target variable. Therefore, we graphically illustrated the overall significance of each attribute in tree models, which are shown in Figure 10. These are of great theoretical value and practical significance for the promotion of low-carbon life, guiding and encouraging low-carbon transportation and forming low-carbon transport behaviors. The age, number of electric bikes, trip duration, trip distance and trip purpose attributes contributed to the modeling.

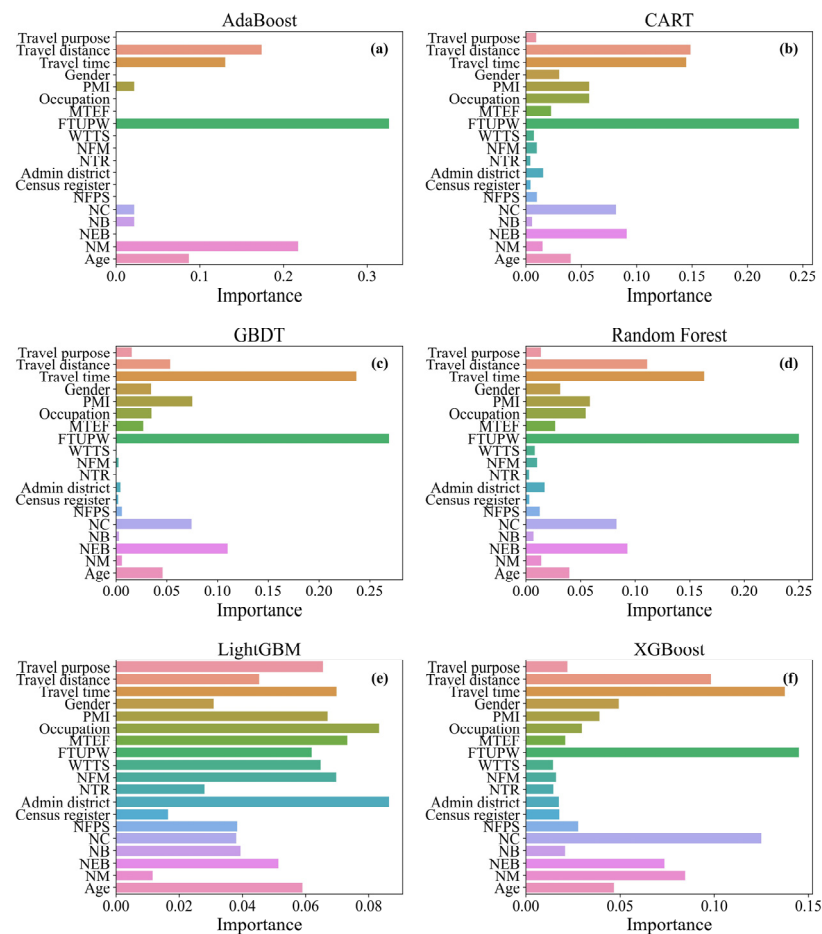


Figure 10. Feature importance for the tree-based classifiers: (a) AdaBoost, (b) CART, (c) GBDT, (d) Random Forest, (e) LightGBM, (f) XGBoost.

7. Conclusions

With the development of the economy and society, the number of urban residents increases sharply, which results in traffic congestion and many other problems. Building an effective multimode transportation system is conducive to alleviating traffic congestion and reducing air pollution. However, it is hard to determine the multimode traffic flow structure due to the complex travel mode choices of travelers. Therefore, understanding and predicting travel mode choices are pivotal to urban traffic management.

Travel mode choice is affected by many factors, such as income, profession, age, gender, etc. In this study, we conducted a quantitative and qualitative analysis to enhance the understanding of the characteristics of travelers and their behaviors using a large amount of survey data from Jinan, China. The results showed that electric bikes have gradually become a dominant travel mode in Jinan city. The main travel purpose of travelers was commuting. There were two peaks (7:00 a.m.–9:00 a.m. and 5:00–7:00 p.m.) for most travel choices except taxis, as people rarely commute by cab but use private cars and electric vehicles instead. People's income could impact the travel mode choice and rich people preferred to use private cars. A stacking machine learning method considering travelers' behavior was introduced. Specifically, a stacking model composed of seven machine learning models was adopted to predict travelers' mode choices. The results indicated that the proposed stacking method achieved accuracy of 0.83, outperforming the traditional MNL mode and nine other machine learning methods. In particular, for the three main travel modes—bus, private car and electric bike—the stacking model showed better performance. We also illustrated the feature importance of the tree-based models.

Travel habit was the dominant factor influencing travelers' mode choice. This study could help in building multimode transportation systems and enhance the travel service quality.

The limitations of this study are as follows. The data used in this study only included one-day trips for each person. In reality, people may use several travel modes on different days. Moreover, some external factors, such as weather, could impact travel mode choices. This study did not consider the impact of these factors. In addition, some trips involve more than one travel mode, and we only used the main travel mode in this study. These research limitations call for future improvements.

Author Contributions: Conceptualization, H.Z.; methodology, H.Z. and Y.L.; writing—original draft preparation, H.Z. and Y.L.; writing—review and editing, L.Z. (Li Zhang) and L.Z. (Lele Zhang). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (grant no. 42001396), the Youth Innovation Team Science and Technology Support Project in Colleges and Universities of Shandong Province (2021KJ058) and the Graduate Education Quality Improvement Plan Program of Shandong Jianzhu University (YZKC202115).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Feng, J.X.; Dijst, M.; Wissink, B.; Prillwitz, J. Changing travel behaviour in urban China: Evidence from Nanjing 2008–2011. *Transp. Policy* **2017**, *53*, 1–10. [[CrossRef](#)]
- Li, C.J.; Zhang, Y.; Chai, Y.W. Do spatial factors outweigh institutional factors? Changes in influencing factors of homework separation from 2007 to 2017 in Beijing. *J. Transp. Geogr.* **2021**, *96*, 103201. [[CrossRef](#)]
- Li, C.Z.; Xiao, W.; Zhang, D.Y.; Ji, Q. Low-carbon transformations of cities: Understanding the demand for dockless bike sharing in China. *Energy Policy* **2021**, *159*, 112631. [[CrossRef](#)]
- Geng, J.C.; Long, R.Y.; Chen, H.; Li, W.B. Exploring the motivation-behavior gap in urban residents' green travel behavior: A theoretical and empirical study. *Resour. Conserv. Recy.* **2017**, *125*, 282–292. [[CrossRef](#)]
- Lin, B.Q.; Wang, X. Does low-carbon travel intention really lead to actual low-carbon travel? Evidence from urban residents in China. *Econ. Anal. Policy* **2021**, *72*, 743–756. [[CrossRef](#)]
- Ding, C.; Wang, D.G.; Liu, C.; Zhang, Y.; Yang, J.W. Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. *Transp. Res. Part A Policy Pract.* **2017**, *100*, 65–80. [[CrossRef](#)]
- Liao, F.X.; Tian, Q.; Arentze, T.; Huang, H.J.; Timmermans, H.J.P. Travel preferences of multimodal transport systems in emerging markets: The case of Beijing. *Transp. Res. Part A Policy Pract.* **2020**, *138*, 250–266. [[CrossRef](#)]
- De Vos, J.; Alemi, F. Are young adults car-loving urbanites? Comparing young and older adults' residential location choice, travel behavior and attitudes. *Transp. Res. Part A Policy Pract.* **2020**, *132*, 986–998. [[CrossRef](#)]
- Shaer, A.; Haghshenas, H. The impact of COVID-19 on older adults' active transportation mode usage in Isfahan. *Iran. J. Transp. Health* **2021**, *23*, 101244. [[CrossRef](#)]
- Mukhamedjanov, A.; Kidokoro, T.; Seta, F.; Yang, Y. Reshaping the concept of transit-oriented development in response to public space overheating near the transit nodes of Tokyo. *Cities* **2021**, *116*, 103240. [[CrossRef](#)]
- Sharif, M.S.; Rahman, M.L. Developing a conceptual framework for an eco-friendly smart urban living. *J. Urban Plan. Dev.* **2022**, *148*, 04022003. [[CrossRef](#)]
- Zheng, H.; Zhang, X.C.; Chen, J.H. Study on customized shuttle transit mode responding to spatiotemporal inhomogeneous demand in super-peak. *Information* **2021**, *12*, 429. [[CrossRef](#)]
- Lopez-Carreiro, I.; Monzon, A.; Lopez-Lambas, M.E. Comparison of the willingness to adopt MaaS in Madrid (Spain) and Randstad (The Netherlands) metropolitan areas. *Transp. Res. Part A Policy Pract.* **2021**, *152*, 275–294. [[CrossRef](#)]
- Ahmad, Z.; Batool, Z.; Starkey, P. Understanding mobility characteristics and needs of older persons in urban Pakistan with respect to use of public transport and self-driving. *J. Transp. Geogr.* **2019**, *74*, 181–190. [[CrossRef](#)]
- Weng, J.C.; Tu, Q.; Yuan, R.L.; Lin, P.F.; Chen, Z.H. Modeling mode choice behaviors for public transport commuters in Beijing. *J. Urban Plan. Dev.* **2018**, *144*, 05018013. [[CrossRef](#)]
- Lodhi, R.H.; Rana, I.A. Mode choice modeling for educational trips in a medium-size city: Case study of Abbottabad city, Pakistan. *J. Urban Plan. Dev.* **2021**, *147*, 05021038. [[CrossRef](#)]
- Lu, X.S.; Liu, T.L.; Huang, H.J. Pricing and mode choice based on nested logit model with trip-chain costs. *Transp. Policy* **2015**, *44*, 76–88. [[CrossRef](#)]

18. McFadden, D.; Train, K. Mixed mnl models for discrete response. *J. Appl. Econom.* **2000**, *15*, 447–470. [[CrossRef](#)]
19. Salas, P.; De la Fuente, R.; Astroza, S.; Carrasco, J.A. A systematic comparative evaluation of machine learning classifiers and discrete choice models for travel mode choice in the presence of response heterogeneity. *Expert Syst. Appl.* **2022**, *193*, 116253. [[CrossRef](#)]
20. Zheng, J.J.; Cheng, Y.; Ma, G.; Han, X.; Yu, L.K. Feasibility analysis of green travel in public transportation: A case study of Wuhan. *Sustainability* **2020**, *12*, 6531. [[CrossRef](#)]
21. Hatamzadeh, Y.; Habibian, M.; Khodaii, A. Commuters' preference to walk: Developing a structural equation model considering current amount of walking and subjective and environmental factors. *J. Urban Plan. Dev.* **2021**, *147*, 04021043. [[CrossRef](#)]
22. Bi, H.; Ye, Z.R.; Hu, L.Y.; Zhu, H. Why they don't choose bus service? Understanding special online car-hailing behavior near bus stops. *Transp. Policy* **2021**, *114*, 280–297. [[CrossRef](#)]
23. Han, Y.; Li, W.Y.; Wei, S.S.; Zhang, T.T. Research on passenger's travel mode choice behavior waiting at bus station based on SEM-Logit integration model. *Sustainability* **2018**, *10*, 1996. [[CrossRef](#)]
24. Si, Y.; Guan, H.Z.; Cui, Y.C. Research on the choice behavior of taxi and express service based on the SEM-Logit model. *Sustainability* **2019**, *11*, 2974. [[CrossRef](#)]
25. Wang, F.; Ross, C.L. Machine learning travel mode choices: Comparing the performance of an extreme gradient boosting model with a multinomial logit model. *Transp. Res. Rec.* **2018**, *2672*, 35–45. [[CrossRef](#)]
26. Cheng, L.; Chen, X.; De Vos, J.; Lai, X.; Witlox, F. Applying a random forest method approach to model travel mode choice behavior. *Travel Behav. Soc.* **2019**, *14*, 1–10. [[CrossRef](#)]
27. Zhao, X.; Yan, X.; Yu, A.; Van Hentenryck, P. Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel Behav. Soc.* **2020**, *20*, 22–35. [[CrossRef](#)]
28. Rafiq, R.; McNally, M.G. A structural analysis of the work tour behavior of transit commuters. *Transp. Res. Part A Policy Pract.* **2022**, *160*, 61–79. [[CrossRef](#)]
29. Molin, E.; Mokhtarian, P.; Kroesen, M. Multimodal travel groups and attitudes: A latent class cluster analysis of Dutch travelers. *Transp. Res. Part A Policy Pract.* **2016**, *83*, 14–29. [[CrossRef](#)]
30. Zhang, Y.S.; Yao, E.J.; Zhang, R.; Xu, H. Analysis of elderly people's travel behaviours during the morning peak hours in the context of the free bus programme in Beijing, China. *J. Transp. Geogr.* **2019**, *76*, 191–199. [[CrossRef](#)]
31. Deng, Y.; Wang, J.X.; Gao, C.; Li, X.H.; Wang, Z.; Li, X.L. Assessing temporal-spatial characteristics of urban travel behaviors from multiday smart-card data. *Phys. A Stat. Mech. Appl.* **2021**, *576*, 126058. [[CrossRef](#)]
32. Zhang, H.; Shi, B.Y.; Zhuge, C.X.; Wang, W. Detecting taxi travel patterns using GPS trajectory data: A case study of Beijing. *KSCE J. Civ. Eng.* **2019**, *23*, 1797–1805. [[CrossRef](#)]
33. Zhang, H.; Zhuge, C.X.; Jia, J.M.; Shi, B.Y.; Wang, W. Green travel mobility of dockless bike-sharing based on trip data in big cities: A spatial network analysis. *J. Clean. Prod.* **2021**, *313*, 127930. [[CrossRef](#)]
34. Zhang, R.; Xie, P.; Wang, C.; Liu, G.Y.; Wan, S.H. Classifying transportation mode and speed from trajectory data via deep multi-scale learning. *Comput. Netw.* **2019**, *162*, 106861. [[CrossRef](#)]
35. Faroqi, H.; Mesbah, M.; Kim, J.; Tavassoli, A. A model for measuring activity similarity between public transit passengers using smart card data. *Travel Behav. Soc.* **2018**, *13*, 11–25. [[CrossRef](#)]
36. Zhang, H.; Zhang, L.L.; Che, F.; Jia, J.M.; Shi, B.Y. Revealing urban traffic demand by constructing dynamic networks with taxi trajectory data. *IEEE Access* **2020**, *8*, 147673–147681. [[CrossRef](#)]
37. Bachir, D.; Khodabandelou, G.; Gauthier, V.; El Yacoubi, M.; Puchinger, J. Inferring dynamic origin-destination flows by transport mode using mobile phone data. *Transp. Res. Part C Emerg. Technol.* **2019**, *101*, 254–275. [[CrossRef](#)]
38. Truong, L.T.; Somenahalli, S.V.C. Exploring frequency of public transport use among older adults: A study in Adelaide, Australia. *Travel Behav. Soc.* **2015**, *2*, 148–155. [[CrossRef](#)]
39. Pettersson, P.; Schmöcker, J.D. Active ageing in developing countries?—Trip generation and tour complexity of older people in Metro Manila. *J. Transp. Geogr.* **2010**, *18*, 613–623. [[CrossRef](#)]
40. Jin, H.; Yu, J. Gender responsiveness in public transit: Evidence from the 2017 US national household travel survey. *J. Urban Plan. Dev.* **2021**, *147*, 04021021. [[CrossRef](#)]
41. Sun, B.D.; Ermagun, A.; Dan, B. Built environmental impacts on commuting mode choice and distance: Evidence from Shanghai. *Transp. Res. Part D Transp. Environ.* **2017**, *52*, 441–453. [[CrossRef](#)]
42. Nguyen, T.M.C.; Kato, H.; Phan, L.B. Is built environment associated with travel mode choice in developing cities? Evidence from Hanoi. *Sustainability* **2020**, *12*, 5773. [[CrossRef](#)]
43. Mao, Z.D.; Ettema, D.; Dijst, M. Commuting trip satisfaction in Beijing: Exploring the influence of multimodal behavior and modal flexibility. *Transp. Res. Part A Policy Pract.* **2016**, *94*, 592–603. [[CrossRef](#)]
44. Ma, S.H.; Yu, Z.L.; Liu, C.Q. Nested logit joint model of travel mode and travel time choice for urban commuting trips in Xi'an, China. *J. Urban Plan. Dev.* **2020**, *146*, 04020020. [[CrossRef](#)]
45. Li, D.W.; Wu, W.T.; Song, Y.C. Comparative study of logit and weibit model in travel mode choice. *IEEE Access* **2020**, *8*, 63452–63461. [[CrossRef](#)]
46. Al-Salih, W.Q.; Esztergar-Kiss, D. Linking mode choice with travel behavior by using logit model based on utility function. *Sustainability* **2021**, *13*, 4332. [[CrossRef](#)]

47. Faber, R.; Merkies, R.; Damen, W.; Oirbans, L.; Massa, D.; Kroesen, M.; Molin, E. The role of travel-related reasons for location choice in residential self-selection. *Travel Behav. Soc.* **2021**, *25*, 120–132. [[CrossRef](#)]
48. Shin, J.; Tilahun, N. The role of residential choice on the travel behavior of young adults. *Transp. Res. Part A Policy Pract.* **2022**, *158*, 62–74. [[CrossRef](#)]
49. Pineda-Jaramillo, J.; Arbelaez-Arenas, O. Assessing the performance of gradient-boosting models for predicting the travel mode choice using household survey data. *J. Urban Plan. Dev.* **2022**, *148*, 04022007. [[CrossRef](#)]
50. Mi, X.Y.; Wang, S.Y.; Shao, C.F.; Zhang, P.; Chen, M.M. Resident travel mode prediction model in Beijing metropolitan area. *PLoS ONE* **2021**, *16*, e0259793. [[CrossRef](#)]
51. Kashifi, M.T.; Jamal, A.; Kashefi, M.S.; Almoshaogeh, M.; Rahman, S.M. Predicting the travel mode choice with interpretable machine learning techniques: A comparative study. *Travel Behav. Soc.* **2022**, *29*, 279–296. [[CrossRef](#)]
52. Xia, Y.T.; Chen, H.F.; Zimmermann, R. A random effect Bayesian Neural Network (RE-BNN) for travel mode choice analysis across multiple regions. *Travel Behav. Soc.* **2023**, *30*, 118–134. [[CrossRef](#)]
53. Elharoun, M.; El-Badawy, S.M.; Shahdah, U.E. Artificial intelligence techniques for predicting individuals' mode choice behavior in Mansoura city, Egypt. *Transp. Res. Rec. J. Transp. Res. Board* **2023**. [[CrossRef](#)]
54. Chen, H.F.; Cheng, Y. Travel mode choice prediction using imbalanced machine learning. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 3795–3808. [[CrossRef](#)]
55. Chorowski, J.; Wang, J.; Zurada, J.M. Review and performance comparison of SVM- and ELM-based classifiers. *Neurocomputing* **2014**, *128*, 507–516. [[CrossRef](#)]
56. Sabzevari, M.; Martinez-Munoz, G.; Suarez, A. Vote-boosting ensembles. *Pattern. Recogn.* **2018**, *83*, 119–133. [[CrossRef](#)]
57. de Ona, R.; Lopez, G.; Rios, F.J.D.D.; Ona, J. Cluster analysis for diminishing heterogeneous opinions of service quality public transport passengers. *Procedia Soc. Behav. Sci.* **2014**, *162*, 459–466. [[CrossRef](#)]
58. Jin, Y.M.; Ye, X.F.; Ye, Q.M.; Wang, T.; Cheng, J.; Yan, X.C. Demand forecasting of online car-hailing with stacking ensemble learning approach and large-scale datasets. *IEEE Access* **2020**, *8*, 199513–199522. [[CrossRef](#)]
59. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurobot.* **2013**, *7*, 21. [[CrossRef](#)]
60. Pirayonesi, S.M.; El-Diraby, T.E. Data analytics in asset management: Cost-effective prediction of the pavement condition index. *J. Infrastruct. Syst.* **2020**, *26*, 04019036. [[CrossRef](#)]
61. Sagi, O.; Rokach, L. Approximating XGBoost with an interpretable decision tree. *Inform. Sci.* **2021**, *572*, 522–542. [[CrossRef](#)]
62. Cheng, L.; De Vos, J.; Shi, K.B.; Yang, M.; Chen, X.W.; Witlox, F. Do residential location effects on travel behavior differ between the elderly and younger adults? *Transp. Res. Part D Transp. Environ.* **2019**, *73*, 367–380. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.