Preprints.org

Article

Travel Behavior and Traffic Safety Perceptions Towards Carsharing and Autonomous Vehicles: A Machine Learning Analysis

Jamil Hamadneh [*] and Noura Hamdan

Article

Travel Behavior and Traffic Safety Perceptions Towards Carsharing and Autonomous Vehicles: A Machine Learning Analysis

Jamil Hamadneh[1,*] and Noura Hamadan [2]

[1] Department of Civil Engineering and Sustainable Structures, Palestine Technical University- Kadoorie, Tulkarem, Palestine

[2] Department of Transport Technology and Economics, Budapest University of Technology and Economics, Budapest, Hungary

* Correspondence: jamil.hamadneh@ptuk.edu.ps

Abstract: The rapid evolution of the transport industry requires a deep understanding of user preferences for emerging mobility solutions, particularly carsharing (CS) and autonomous vehicles (AVs). This study employs machine learning techniques to model transport mode choice, with a focus on traffic safety perceptions of people towards CS and privately shared autonomous vehicles (PSAVs). A stated preference (SP) survey is conducted to collect data on travel behavior, incorporating key attributes such as trip time, trip cost, waiting and walking time, privacy, cybersecurity, and surveillance concerns. Sociodemographic factors, such as income, gender, education, employment status, and trip purpose, are also examined. Three gradient boosting models—CatBoost, XGBoost, and LightGBM are applied to classify user choices. The performance of models is evaluated using accuracy, precision, and F1-score. The XGBoost demonstrates the highest accuracy (77.174%) and effectively captures the complexity of mode choice behavior. The results indicate that CS users are easily classified, while PSAV users present greater classification challenges due to variations in safety perceptions and technological acceptance. From a traffic safety perspective, the results emphasize that companionship, comfort, privacy, cybersecurity, safety in using CS and PSAV, and surveillance significantly influence CS and PSAV adoption, which leads to the importance of trust in adopting AVs. The findings suggest ensuring public trust through robust safety regulations and transparent data security policies. Furthermore, the envisaged benefits of shared autonomous mobility are alleviating congestion and promoting sustainability.

Keywords: autonomous vehicle; carsharing; transport mode choice; machine learning

1. Introduction

The transport sector has undergone a significant transformation in recent decades, driven by rapid technological advancements that have introduced novel mobility solutions [1,2]. Among these innovations, carsharing (CS) services and autonomous vehicles (AVs) have gained substantial attention due to their potential to reshape urban mobility [3,4]. These emerging transport mode services offer alternatives to traditional car ownership by enhancing efficiency, safety, sustainability, and flexibility [5]. The CS and the AVs are anticipated to be essential in creating sustainable and intelligent transportation systems as urbanization increases, traffic congestion grows, and environmental concerns rise [6-8]. Furthermore, one of the most important factors impacting AV user adoption is safety concerns, which continue to be a major factor in travel behavior and decision-making [9,10].

CS services provide users with on-demand vehicle access, thereby reducing the number of private vehicles on the road, alleviating congestion, and mitigating environmental impacts, especially

at city centers [11,12]. Meanwhile, AVs have the potential to significantly increase road safety, fuel economy, and traffic efficiency [13-15]. By minimizing human error, one of the leading causes of road accidents, autonomous driving technology has the potential to significantly reduce traffic fatalities and injuries [16]. However, the perceived safety of AVs remains a major barrier to widespread adoption, as users express concerns regarding system reliability, cybersecurity threats, and the ability of automated systems to handle complex driving scenarios [17-20].

Numerous studies have examined the factors influencing the adoption of CS and AVs, emphasizing elements such as convenience, cost, environmental awareness, and trust in technology [21-26]. The interaction between safety perceptions, behavioral tendencies, and demographic characteristics in shaping mode choice remains underexplored. Conventional discrete choice models have provided valuable insights into user preferences [27-29]. Conventional discrete choice models often fail to capture the complexity of decision-making in the evolving transport landscape. In order to overcome these constraints, machine learning (ML) has become a powerful analytical technique that makes it possible to analyze large, diverse datasets and find hidden patterns and interdependencies among the variables driving transport decisions [30-32].

Previous research has demonstrated the effectiveness of machine learning models, ranging from traditional classifiers such as logistic regression and decision trees to advanced algorithms like Random Forest, Gradient Boosting, and Neural Networks, in improving prediction accuracy for mode choice behavior [31,33,34]. However, studies applying machine learning to analyze the safety-related concerns influencing the adoption of CS and PSAVs services remain limited [13,35]. It is noted that PSAVs represent full automation, allowing users to experience autonomous mobility, whereas traditional CS services rely entirely on human drivers. The distinction between these two services requires a comparative analysis to understand how safety perceptions influence user preferences. Furthermore, studying the preferences of people towards these two services.

The added value of this research to the literature is summarized in applying three high-performance machine learning models—CatBoost, XGBoost, and LightGBM—to classify user choices between CS and PSAVs. The study aims to identify the key determinants influencing transport mode selection, assess the role of safety-related concerns in shaping user preferences, and evaluate the comparative performance of machine learning algorithms in predicting travel behavior. By integrating demographic, behavioral, and trip-related variables, this research seeks to provide actionable insights for policymakers, urban planners, vehicle manufacturers, and mobility service providers. This research answers the following questions:

- What are the factors that impact the adoption of CS and PSAVs?
- Do machine learning models predict whether a traveler is likely to adopt CS or PSAVs based on safety perceptions?
- How do travelers inside cities perceive the safety of CS and PSAVs?
- Do people's selection of PSAVs and CS vary by demographic variables?

This research article is organized as follows: Section 1 presents the introduction, which includes previous research gaps, current research contributions, and research aims. The literature review is presented in Section 2, which presents previous related research. The methodology explains the tools and techniques used, is presented in Section 3. Section 4 presents the results and discussions. Section 5 presents the study's summary and conclusions.

2. Literature Review

In recent years, CS and AV services have attracted a lot of attention as a practical way to address issues with transport in cities, such as traffic, parking restrictions, and environmental concerns [36,37]. Safety considerations play a crucial role in shaping travel behavior, particularly in emerging mobility services such as CS and PSAVs [38]. Extensive research has explored the impact of safety perceptions on transport mode choices, addressing concerns related to technological reliability, accident risks, cybersecurity, and personal security [39-41]. This section provides an overview of previous relevant studies on traveler preferences towards CS and AVs, and safety-related factors

influencing CS and AV adoption, while also examining the role of machine learning in predicting travel behavior.

A study by Kyriakidis, et al. [42] was conducted to analyze public perceptions of AVs based on survey responses from eight European countries. The authors found that safety remains a primary determinant of AV adoption, with demographic factors such as age, gender, education, and household size influencing willingness to use AVs. Vulnerable road users, including the elderly and individuals with disabilities, expressed a preference for human supervision in AVs, highlighting broader concerns related to reliability, cost, and driving experience. These insights suggest that regulatory frameworks should consider both safety and user comfort in AV implementation. A study by Stoiber, et al. [43] explores user preferences for pooled AVs through an online choice experiment with 709 participants in Switzerland. The study assesses both short- and long-term mobility decisions based on a scenario of full AV market penetration. The study results indicate that 61% of respondents preferred shared AVs over private autonomous cars, reinforcing the potential of pooled AV services to reduce private vehicle dependence. Additionally, integrated measures addressing cost, travel time, and comfort are identified as critical factors in promoting shared mobility solutions [43]. Zhou, et al. [44] conducted a study to examine consumer preferences toward CS and the potential adoption of shared automated vehicles (SAVs) through a stated preference survey in Australia. A mixed logit model is applied, and the study reveals substantial preference heterogeneity, with prior experience in car-sharing increasing multimodal travel choices while reducing private vehicle reliance. The authors show that elderly people, women, and non-drivers—who are generally viewed as major SAV beneficiaries—show lower levels of acceptability, underlining potential challenges to broad adoption and the importance of focused policy measures.

In a study by Kolarova, et al. [45], the travelers are more likely to use AVs than conventional cars. The AVs offer wider choices for riders on board, such as engaging in active and passive activities where driving negative sides are released, as shown in studies of Hamadneh and Esztergár-Kiss [5] and Hamadneh and Esztergár-Kiss [46]. Hao and Yamamoto [37] find that car ownerships are less likely to use CS than people who do not own cars in urban areas. Moreover, they find that the pricing strategies in urban areas are a factor that impacts the use of CS where people park their cars outside of the city and use CS or other transport modes. The AVs are still not on the market, researchers use mathematical models based on questionnaires and surveys to understand the behavior of people towards AVs as part of a transport system. It is found that personal experiences impact the adoption of CS and AVs, as stated by Müller [47]. Schoettle and Sivak [48] examine how travelers in the USA, UK, and Australia deal with the availability of AVs in the market. The authors find that preference towards AVs changes across gender, such as women are more willing to use AVs than men. Moreover, a study by Howard and Dai [49] states that males who are highly educated, own luxury cars, and are high-income earners are more willing to use AVs compared to other groups. Additionally, studies by Bansal, Kockelman [50], and Megens, et al. [50] state that males and high-income people are more likely to use SAVs. Women hesitate to buy AVs due to safety factors, therefore, they are less likely to pay more money for automation, as stated by Payre, et al. [51].

Scholars use mixed logit models to understand the expected behavior of people when AVs are in the market. Yan, et al. [52] finds that AVs are accepted by users if trip time, trip cost, and waiting time are competent. In Japan, a study by Das, et al. [53] shows that around 20-30% of trips are willing to be changed when AVs are in the market, based on the results of a nested logit model. Moreover, the author states that AVs are impacted by job types, such as part-time workers are more willing to use AVs. On the other hand, CS is considered an option that attracts travelers in urban areas. The CS is considered a cost-effective option in cities more than privately owned cars, where the cost of parking and using infrastructure is eliminated [54,55]. Fornahl and Hülsmann [56] show that CS is a solution to traffic congestion in city centers where people can rent a car for a short period,   parking is utilized, and traffic congestion is alleviated. Efthymiou, et al. [57] shows that CS users are among the youngest and educated ones. While Shaheen, et al. [58] states that CS are used by students who

own a driving license and have low income. Males are the main users of CS based on studies of Ciari, et al. [59] and Perboli, et al. [60] in Italy and Switzerland.

Lee [61] analyzes the changing dynamics of transportation mode choice in the AV era through a combination of discrete choice modeling (DCM) and machine learning (ML) techniques. A stated choice experiment in the U.S. reveals that AV market shares are influenced by a range of socio-demographic and behavioral factors. The study utilizes stochastic gradient boosting to enhance feature interpretability, uncovering non-linear relationships between user characteristics and mode choice. Additionally, methodological limitations in ML-based mode choice modeling are critically assessed, highlighting areas for future refinement. Pineda-Jaramillo, et al. [62] compares traditional multinomial logit models with ML approaches in travel mode choice prediction. In their study, based on household survey data from the Aburrá Valley, Colombia, they found that an optimized gradient boosting model outperformed both logit and random forest models. Key determinants of mode choice included travel time, parking availability, vehicle ownership, age, and gender, demonstrating the potential of machine learning as a policy tool for promoting sustainable transport options.

In sustainability, AVs and CS participate in the reduction of greenhouse gas emissions as well as promote eco-friendly transport modes, as stated by Hao and Yamamoto [37]. Particularly, the transport sector is the cause of almost one-third of the greenhouse gas emissions in the world, as stated by the WHO, because of the use of internal combustion engines, traffic congestion, and the rise in car ownership worldwide. Furthermore, using CS and AVs, which are on-demand transport systems, can help alleviate the negative impact of conventional transport modes as well as promote sustainable travel behavior in cities [36]. CS and AVs' influences are realized in the reduction of the pollution resulting from alleviating traffic congestion, increasing the efficiency of fuel, electric vehicles, less parking, development of engines, and optimizing the fleet size of AVs and CS [63-65].

In summary, this research applies a novel approach where CS and PSAVs are studied based on the safety perceptions of travelers in urban areas, considering the main trip of travelers in Budapest, Hungary. Moreover, the travelers' behaviors towards CS and PSAV are examined. This study fills the gap in the literature where scarcity of research that finds the impact of certain factors on the choice of CS and PSAV services using machine learning. This study employs machine learning techniques—CatBoost, XGBoost, and LightGBM—to analyze the factors influencing mode choice between CS and PSAVs. By integrating demographic, behavioral, and safety-related variables, this approach aims to provide a comprehensive understanding of the complex decision-making processes involved in adopting these emerging mobility solutions.

3. Methodology

This section presents the methods used in conducting the research. The methodology is divided into two subsections as shown in Figure 1, as follows: data collection and data analysis.
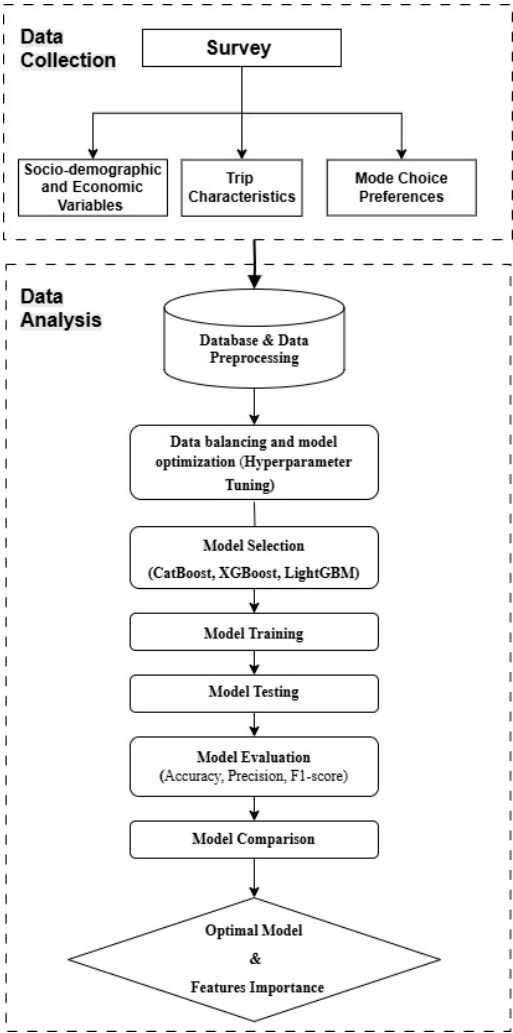
Figure 1. Research approach.

3.1. Data Collection

The dataset used in this study is taken from a stated preferences (SP) survey conducted online in 2021 in Budapest, Hungary. A total of 1840 responses are used in this research. The dataset encompasses both numerical and categorical variables, with columns that describe user preferences, transport mode choices, and socio-demographic features. The SP survey contains three parts: sociodemographic variables, main trip characteristics, preferred factors during travel, and transport mode alternatives.

The collected data includes the following variables that are grouped in sections, as shown in Table 1. Sociodemographic variables, trip characteristics, preferences towards travel, and transport alternatives based on specific attributes are collected.

Table 1. The SP survey contents.

Section	Features
Sociodemographic variables	Gender, age, income, car ownership, job, education
Main trip characteristics	Most frequent transport mode, trip length, trip purpose assuming using CS and PSAV
Preferred factors during travel	Waiting time, transport cost, comfort, reliability, safety, privacy, traffic congestion, companion onboard, cybersecurity
Transport mode choices (i.e., CS and PSAV)	Trip time, trip cost, time to start the trip, availability of onboard camera (i.e., surveillance control)

3.1.1. Descriptive Statistics

The number of received responses is 1840 after filtering the data. The sociodemographic data include income, job, educational level, age, gender, car ownership, and educational levels. 9.30% of age groups are 15-24 years old, 84.7% are 25-54 years old, 4.5% are 55-64 years old, and 1.6% are 65 years old and above. People with low income represent 27.20% of the respondents, while middle-income respondents are 29.40%. The proportion of high-income respondents is 30.70%. Respondents who have work are 70.70% of the respondents are working either full time, part time, or self-employed. Students represent 13.10% of the survey participants, retired people are 3.80% of the participants, unemployed resp are 8.00%, and other categories are 4.20% of the participants. Female participants are 44.70%, and males are 55.30%. In car ownership, 58.8% of participants own cars. 38.00% of participants hold high graduate studies, while 42.20% hold undergraduate studies. 11.20% are high school holders, and only 8.6% hold different education levels.

Table 2 shows how travelers' preferences change across age groups. Age group of 25-34 years is more likely to use CS and PSAV than other groups.

Table 2. Travelers' choices across age groups.

	15-24	25-34	35-44	45-54	55-64	65+
CS	9.5%	55.9%	21.9%	6.0%	4.6%	2.1%
PSAV	8.9%	53.8%	22.4%	9.6%	4.2%	1.0%

Table 3 shows the number of responses towards CS and PSAV per trip purpose. For example, 8.64% of responses demonstrate that educational trips are more preferred to be made by CS than PSAV, while 8.42% of responses demonstrate that educational trips are more preferred to be made by PSAV than CS. It is noted from Table 3 that travelers have small differences in using CS and PSAV in their travel to educational trips. On the contrary, there are differences in using CS and PSAV to and from home, shopping, leisure, and work trips. Shopping and work trips are preferred to be made by PSAV, while home and leisure trips are preferred to be made by CS.

Table 3. Trip purpose by CS and PSAV.

Transport Mode	Trip Purpose				
	Education	Home	Shopping	Leisure or others	Work
CS	8.64%	7.12%	13.59%	16.20%	54.45%
PSAV	8.42%	5.82%	26.25%	13.48%	46.03%

Figures 2 and 3 show the distribution of responses across gender types. It is shown that females are more likely than males to use PSAV and CS for educational and entertainment trips than males.
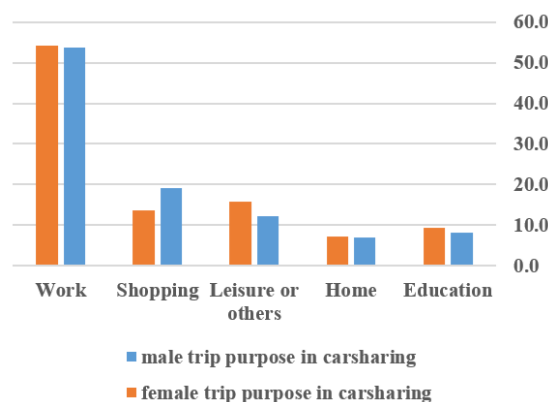

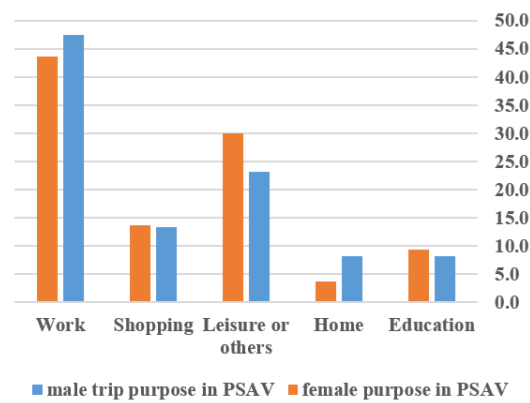**Figure 2.** Trip purpose by CS across gender.

Figure 3. Trip purpose by PSAV across gender.

Figure 4 shows the traveler's transport modes and income classes. It is shown that various transport modes are used across different classes of income. For example, a car as a passenger is used by high-income people, a car as a driver is mainly used by high- and middle-income classes, a bike is used by high-income people, public transport is mainly used by low-income people, and walking is used by high- and middle-income people.
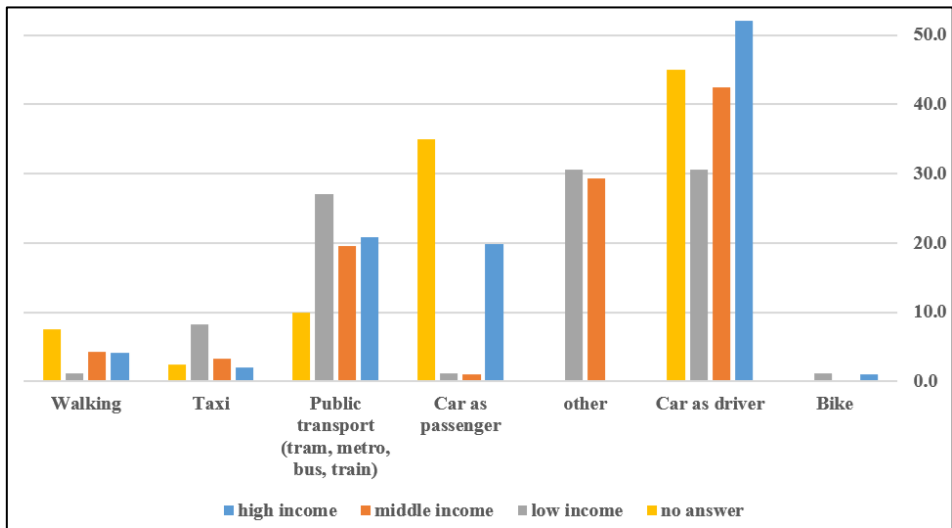

Figure 4. Income classes of respondents across travelers' current transport mode.

Figure 5 presents transport mode margins where travelers select to conduct their main trips in urban areas. This share demonstrates that people are more willing to choose CS over PSAV.
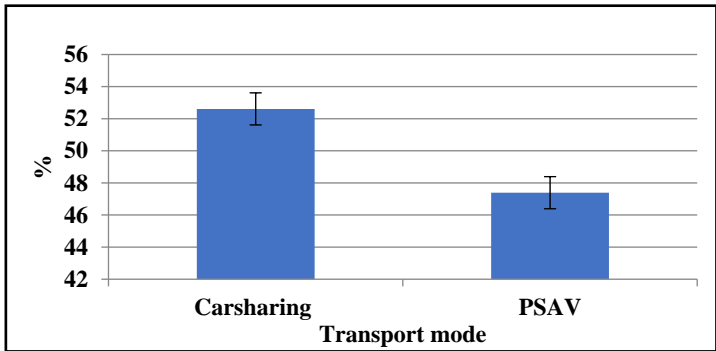

Figure 5. Transport mode shares by travelers.

Figure 6 presents the preferences of travelers towards CS and PSAV regarding whether the camera is installed on board or not. The selection of travelers demonstrates that travelers choose CS over PSAV when a surveillance camera is installed on board. On the other hand, there are still considerable shares of using PSAVs in their travel to their main trips when camera is not installed on board. The factors that impact the preferences of people are explained by the result of machine learning.
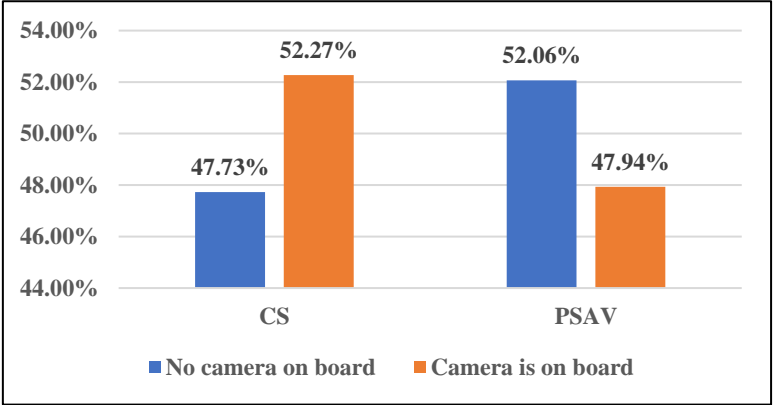

Figure 6. Transport mode's shares by travelers by the availability of cameras on board.

Figures 7 and 8 shows the preferences of travelers on their main trip when they choose to travel by CS and PSAV, respectively.
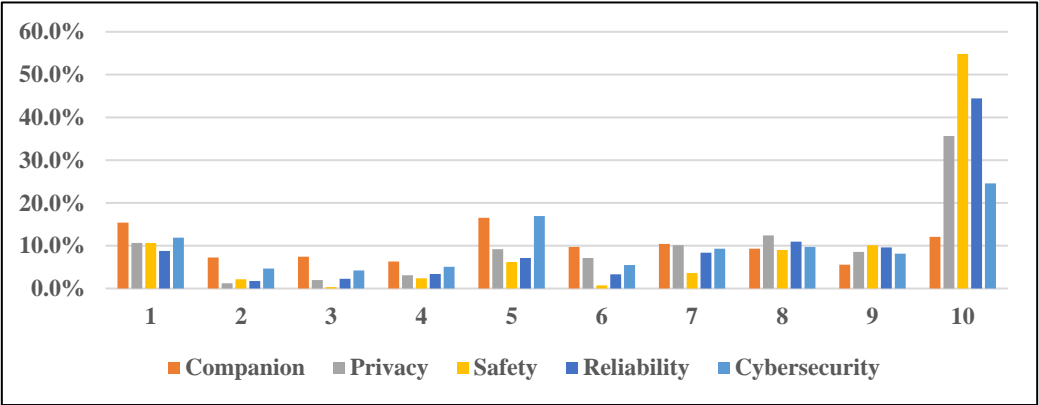

Figure 7. CS and safety preferences of travelers in urban areas.
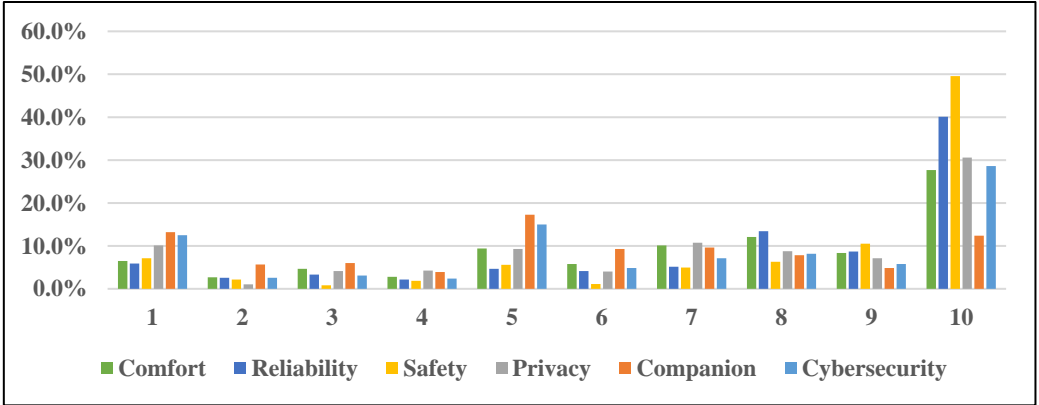

Figure 8. PSAV and safety preferences of travelers in urban areas.

Figure 9 shows the preferences of people driving or riding to their main trips in urban areas. The classifications are as follows: 1- Extremely not important, 2- Moderately not important, 3- Little not

important, 4- Mildly not important, 5- Partially not important, 6-Partially important, 7- Mildly important, 8- Little more important, 9- Moderately important, and 10- Extremely important. It is shown from Figure 9 that companion is the least important factor based on the number of responses (48.1% of responses based on scale 6 and more), while safety occupies the highest level of importance, followed by reliability (79.5% and 78.2% of responses, respectively, based on scale 6 and more). Privacy is the third important factor (71.2% of responses based on scale 6 and more), and comfort and cybersecurity are nearly equally important (69.2% and 59.0% of responses based on scale 6 and more).
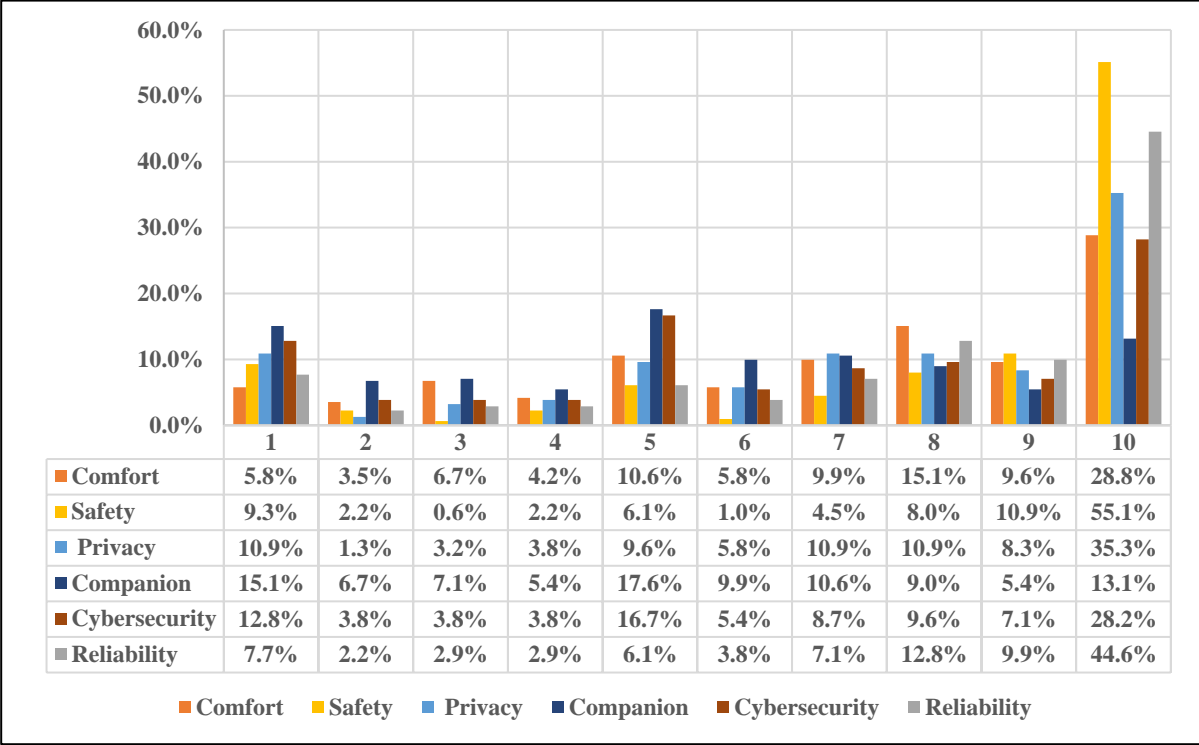

	1	2	3	4	5	6	7	8	9	10
Comfort	5.8%	3.5%	6.7%	4.2%	10.6%	5.8%	9.9%	15.1%	9.6%	28.8%
Safety	9.3%	2.2%	0.6%	2.2%	6.1%	1.0%	4.5%	8.0%	10.9%	55.1%
Privacy	10.9%	1.3%	3.2%	3.8%	9.6%	5.8%	10.9%	10.9%	8.3%	35.3%
Companion	15.1%	6.7%	7.1%	5.4%	17.6%	9.9%	10.6%	9.0%	5.4%	13.1%
Cybersecurity	12.8%	3.8%	3.8%	3.8%	16.7%	5.4%	8.7%	9.6%	7.1%	28.2%
Reliability	7.7%	2.2%	2.9%	2.9%	6.1%	3.8%	7.1%	12.8%	9.9%	44.6%

Figure 9. Safety preferences of travelers on their main trip in urban areas.

Figure 10 demonstrates the differences in preferences across CS and PSAV. People show higher importance of CS safety factors (physical safety, reliability, comfort, privacy, companion, and cybersecurity).
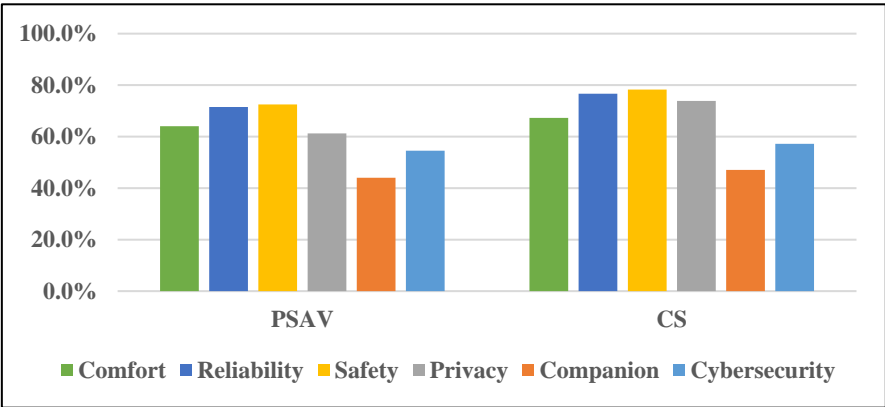

Figure 10. Safety preferences of travelers across CS and PSAV.

3.1.2. Data Processing

In the data collection, categorical variables are encoded using Label Encoding to transform non-numeric features into a numeric form that is consistent with the input format of machine learning models. Whether a passenger chooses CS or PSAV is the desired variable. As seen in

Figure 11, a correlation matrix is created to examine the connections between the attributes and the target variable. The correlation matrix shows the strength and direction of correlations between variables. This analysis provides important insights into how different factors, such as trip cost, waiting time, and sociodemographic variables (e.g., income and education), are interrelated. Therefore, key features are identified that strongly influence user preferences and subsequently focus on the most relevant predictors in the ML models.
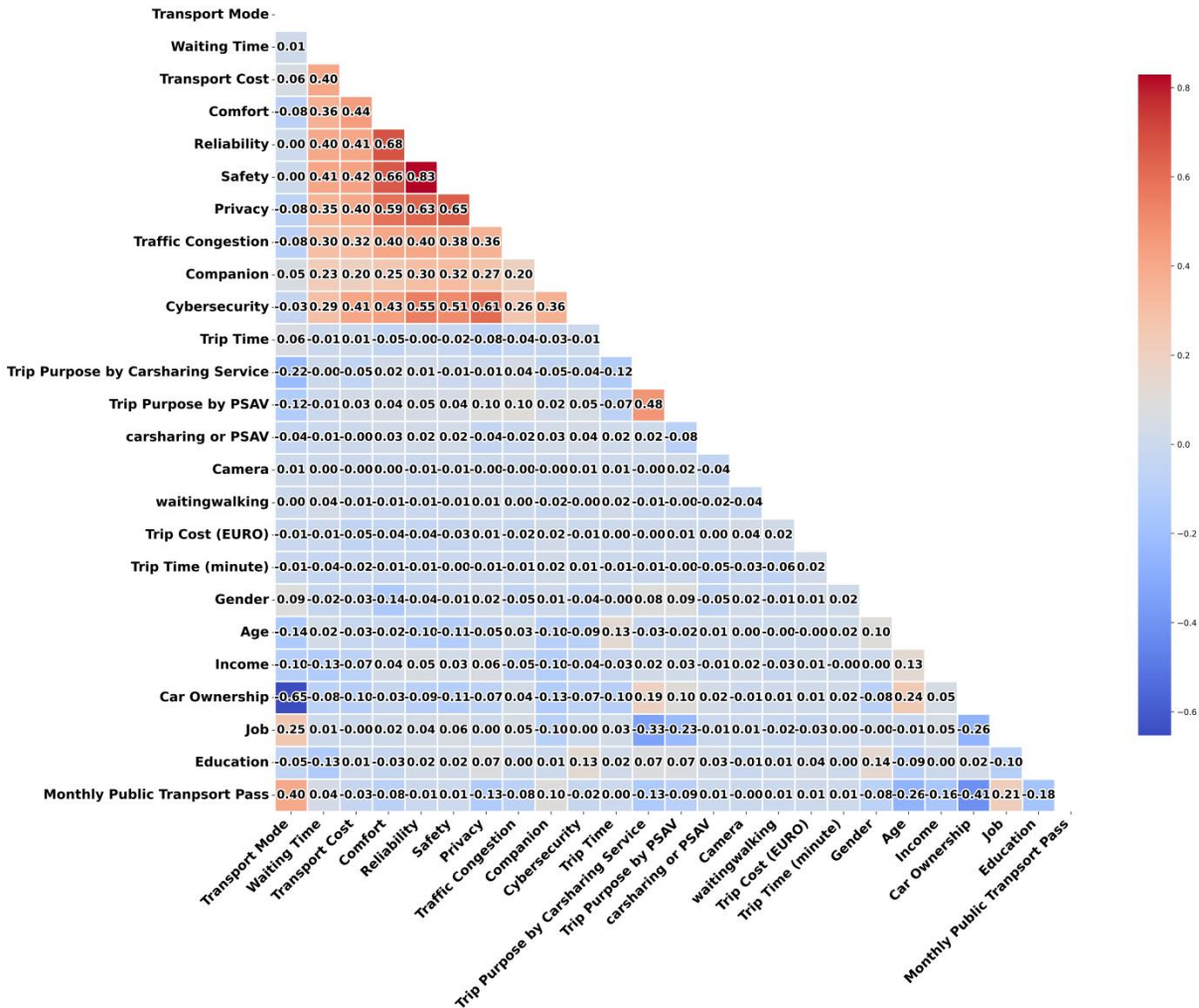

Figure 11. Correlation matrix.

3.1.3. Data Balancing and Model Optimization

The Synthetic Minority Over-Sampling Technique (SMOTE) is applied. SMOTE is used to alleviate the impact of class imbalance observed within the dataset [66]. SMOTE improves class distribution by generating synthetic instances of the minority class. SMOTE prevents bias toward the majority class in the predictive model [66]. This method improves model stability and generalizability by ensuring a more equitable representation of both classes within the training set [67]. Furthermore, optimizing the hyperparameters of XGBoost, LightGBM, and CatBoost is a critical step in enhancing model performance [68]. Various hyperparameter tuning methodologies exist, including Grid Search, Random Search, Evolutionary Algorithms, and Bayesian Optimization [69]. Grid Search and Evolutionary Algorithms perform an exhaustive search across the parameter space, where their computational cost can be prohibitive. Random Search can produce a well-approximated configuration but does not ensure the identification of a globally optimal solution, while being computationally more efficient [70]. In this study, Grid Search, including multiple trials, is utilized to thoroughly optimize and refine the booster parameters of XGBoost, LightGBM, and CatBoost

hyperparameters through intently iterating through key elements within the training dataset. Such systematic tuning improves model generalization, predictive accuracy, and computational efficiency. Together with SMOTE for class balancing and fine-tuning of hyperparameters, this methodology guarantees the resulting predictive models are robust, well-calibrated, high-performing, and significantly reduce the risk of overfitting and underfitting.

3.2. Data Analysis: Model Selection and Training

Data analysis includes three ML classifiers, model selection, and model training to be used in predicting the travelers' preferences towards CS and PSAVs. The models are CatBoost, XGBoost, and LightGBM.

3.2.1. CatBoost

CatBoost is a gradient boosting algorithm based on decision trees, specifically designed to efficiently handle categorical features while mitigating prediction bias. One of its primary advantages is its ability to process categorical variables directly without extensive preprocessing, thus improving both prediction accuracy and model generalization [71].

Categorical features are discrete variables, typically represented as strings, where each unique value corresponds to a specific category. These features cannot be directly utilized as numerical inputs and must undergo a transformation process. CatBoost addresses this by applying an encoding technique that involves scrambling the order of the dataset, denoted as $D=[(x_i,y_i)]i=1,...,n$. The scrambled sequence, $\sigma=(\sigma_1,\ldots,\sigma_n)$, is then used to compute the categorical feature value iteratively. The calculation follows Equation 1:

$$\sigma_{z,k} = \frac{\sum_{j=1}^{z-1}[x_{\sigma_{j,k}} = x_{\sigma_{i,k}}] \cdot Y_{\sigma_{j,k}} + \alpha \cdot z}{\sum_{j=1}^{z-1}[x_{\sigma_{j,k}} = x_{\sigma_{i,k}}] + \alpha} \tag{1}$$

where z represents the prior term, and $\alpha > 0$ is the weight coefficient of the prior term. The inclusion of this prior term helps to reduce the noise associated with low-frequency categorical features. In regression problems, this prior term corresponds to the mean value of the dataset labels.

One of the challenges in gradient boosting decision trees (GBDT) is the presence of gradient bias and overfitting, which arise from using the same data points for gradient estimation. To overcome this, CatBoost employs the Ordered Boosting method, which transforms the gradient estimation process from biased to unbiased. This method first generates a random permutation $\sigma=[1,n]$ to sort the original dataset, initializing n different models $M_1,M_2,\ldots, M_n$. Each model M_i is trained using only the top i samples of the permutation, ensuring that the unbiased gradient estimate of the jth sample is obtained by model M_{j-1} at each iteration. This approach effectively reduces overfitting and enhances model robustness[72].

3.2.2. XGBoost

It represents an ensemble of decision trees built upon the principles of gradient boosting, with a primary focus on scalability[70]. Much like traditional gradient boosting, XGBoost constructs an iterative expansion of the objective function by minimizing a specific loss function. Notably, XGBoost exclusively employs decision trees as base classifiers, employing a modified loss function to regulate the complexity of these trees. These gradient-boosted decision trees, also known as ensemble techniques, have proven to be highly effective in various fields. Within the context of XGBoost, the formulation involving K tree functions is expressed as in Equation (2).

$$\hat{q}_i^{(t)} = \sum_{k=1}^{t} f_k(p_i) = \hat{\boldsymbol{q}}_i^{(t-1)} + \boldsymbol{f_t}(\boldsymbol{p_i}) \tag{2}$$

Where:
- $\hat{q}_i^{(t)}$ signifies the estimated crash severity after the iterations,

- k represents the number of additive trees,
- t denotes the number of iterations,
- $f_k(p_i)$ corresponds to the kth tree function for variables p_i,
- $\hat{q}_i^{(t-1)}$ represents the predicted response value for the final iteration,
- $f_t(p_i)$ characterizes the tree function for the ith iteration.

The objective function for minimizing the loss $l(q_i, \hat{q}_i)$ is structured as in Equations (3) and (4):

$$Obj = \sum_{k=1}^{n} l(q_i, \hat{q}_i) + \sum_{k=1}^{t} \Omega(f_k) \tag{3}$$

$$\Omega(f_t) = \gamma T + \frac{1}{2}\gamma \sum_{j=1}^{T} \omega_j^2 \tag{4}$$

Here, $\Omega(\boldsymbol{f_t})$ acts as the regularization term, preventing overfitting and reducing complexity. T signifies the number of leaves, $\boldsymbol{\omega_j^2}$ represents the *L₂* norm of the *jᵗʰ* leaf scores, and n reflects the total number of crashes in the sample data. This comprehensive approach combines the foundational concepts of gradient boosting and the specialized attributes of XGBoost to create a powerful tool for enhancing prediction accuracy and scalability in machine learning applications.

3.2.3. LightGBM

This study capitalizes LightGBM, a renowned gradient boosting algorithm recognized for its exceptional speed and efficiency. LightGBM, introduced by Ke, et al. [73]_ENREF_29, is a comprehensive library that encompasses gradient boosting and introduces multiple innovative features. LightGBM operates as a supervised model, striving to determine an approximate function *f*(x)* for a given dataset *D= [(xᵢ, yᵢ)]*. This function aims to minimize the loss function *L (y, f(x))*, as depicted in the Equation (5).

$$f = arg\ \min_{f}[E_{y,X}L(y, f(x))] \tag{5}$$

It leverages regression trees, denoted as *fₜ(x)*, combining them based on certain rules or probabilities to approximate *f*(x)*, as articulated in Equation (6).

$$f^*(x) = \sum_{t=1}^{N} \omega_t \times f_t(x) \tag{6}$$

While LightGBM excels at modeling the statistical characteristics of samples for accurate classification and regression, it encounters challenges with unbalanced datasets, which are frequently encountered in intrusion detection scenarios.

3.2.4. Performance Evaluation Metrics

In this study, the performance of three distinct machine learning models: CatBoost, XGBoost, LightGBM is rigorously assessed. Each model is trained using a training set, and predictions are made on the test set. Model performance is evaluated by using various metrics, including accuracy, precision, and F1-score. The accuracy score provides the proportion of correct predictions, while precision and F1-score offer a more detailed view of how well each model performs across different classes [74]. These metrics are computed for each classifier.

4. Results and Discussion

This section presents the findings of the machine learning models applied to predict user mode choice between CS and PSAVs, focusing on traffic safety perceptions and behavioral-demographic influences. The performance evaluation of the models, classification metrics, and feature importance is discussed in detail.

4.1. Model Performance and Comparative Analysis

The predictive performance of the applied machine learning models—XGBoost, CatBoost, and LightGBM—is evaluated using test accuracy, precision, and F1-score. Table 2 presents a comparative summary of the models' performance metrics.

Table 4. Model Performance Evaluation.

	XGBoost	**CatBoost**	**LightGBM**
Accuracy	0.77173913	0.763586957	0.730978261
F1-Score	0.771230089	0.763255229	0.730119138
Precision	0.771869087	0.763478721	0.731100159

Among the tested models, XGBoost achieved the highest predictive performance, with an accuracy of 77.17%, followed by CatBoost (76.36%) and LightGBM (73.10%). The high accuracy and precision of XGBoost suggest its superior ability to capture the complex relationships within the dataset. Figure 12 illustrates the comparative performance of the models, demonstrating the slight variation in classification accuracy. While all models performed competitively, the relatively lower performance of LightGBM may indicate its reduced capacity to generalize under the given data distribution.
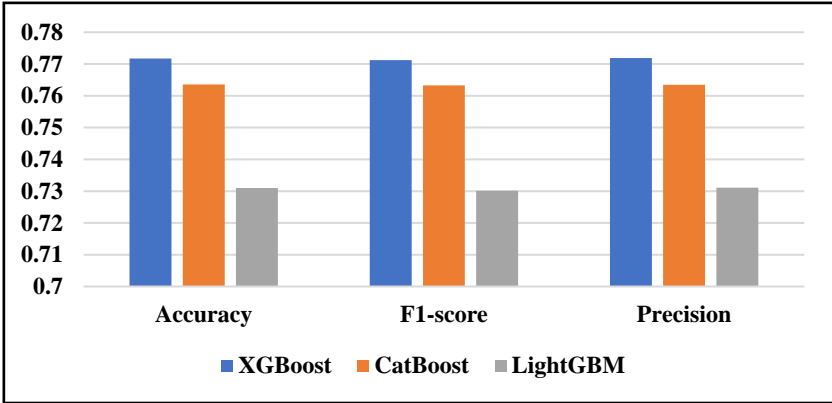

Figure 12. Performance of models.

4.2. Classification Metrics

A detailed evaluation of the models' classification performance is provided in terms of precision, recall, and F1-score for both CS and PSAV. The results reveal that all models exhibit higher recall for CS users than for PSAV users, suggesting a stronger ability to correctly classify CS users.

Table 5. Evaluation of The Models' Classification Performance.

	XGBoost		**LightGBM**		**CatBoost**	
	CS	**PSAV**	**CS**	**PSAV**	**CS**	**PSAV**
Precision	0.77	0.77	0.73	0.73	0.77	0.76
Recall	0.81	0.73	0.78	0.68	0.79	0.73
F1-score	0.79	0.75	0.75	0.70	0.78	0.74

XGBoost achieves a precision of 0.77 for both CS and PSAV, with an F1-score of 0.79 for CS and 0.75 for PSAV, indicating balanced classification performance. CatBoost performs slightly low, with an F1-score of 0.78 for CS and 0.74 for PSAV, showing a small reduction in predictive power compared to XGBoost. LightGBM exhibits the lowest recall for PSAV (0.68), which may imply a greater tendency to misclassify PSAV users compared to the other models.

The classification results indicate that while all models perform well, XGBoost consistently outperforms the others in detecting and correctly classifying users of both transport modes.

4.3. Feature Importance Analysis

To understand the key factors influencing users' mode choice decisions (refer to Figures 5 and 6), feature importance analysis is conducted using the XGBoost model, which demonstrates the highest predictive performance. Figure 13 presents the ranked importance of input features, highlighting the most influential variables in mode choice prediction.

The results provide valuable insights into the behavioral and demographic factors influencing user preferences in CS and PSAVs. The superior performance of XGBoost suggests that decision tree-based ensemble methods are well-suited for capturing the complexity of travel behavior. The classification results indicate that CS users are more accurately identified, PSAV users present greater classification challenges, likely due to variations in safety perceptions and technological acceptance.

The influential factors are divided into three groups for ease of understanding: group 1 is traveling variables, group 2 is sociodemographic variables, and group 3 is safety variables.

Group 1 (traveling variables) includes, trip time, waiting time for PSAV or walking time to and from CS, trip purpose by CS, trip purpose by PSAV, trip cost, traffic congestion, waiting time, current transport mode, usual transport cost, usual main trip time, and camera on board of CS and PSAV. From **Error! Reference source not found.**, the highest influential factor is the trip time using CS and PSAV (6.68%), the duration of the trip plays a critical role in users' mode choice. The third influential factor in the adoption of CS and PSAV is walking and waiting time (6.12%). It is shown from the results that trip cost (4.66%) is the sixth influential factor that impacts the use of CS and PSAV after the trip purpose of using CS (5.09%). While trip purpose by PSAV is the tenth influential factor, which means the people are still studying using PSAV based on destinations. Traffic congestion is the ninth influential factor that people consider when using CS and PSAV. In addition, availability of a camera on board of CS and PSAV slightly impacts the use of CS and PSAVs compared to the other 24 factors.

Moreover, the people consider their current characteristics when they decide to choose CS and PSAV, such as waiting time, the type of transport mode that a traveler uses impact using PSAV and CS, and this applies to the amount of money a traveler pays to use current transport mode as well as the current trip time using current transport mode. This leads to the conclusion, the current travel patterns influence the use of other transport modes in the future, such as CS and PSAVs. The importance of these variables occupies almost the least importance compared to other factors, as shown in Figure 13.

Group 2 (sociodemographic variables) includes car ownership, monthly public transport pass ownership, gender, age, income, education, and job variables. The car ownership (6.12%) is the second influential factor; users with private vehicles may be less inclined to use CS or PSAV. Users of public transport mode are more likely to use shared mobility, as demonstrated in this study, where monthly public transport pass ownership factor occupies the fourth important factor in influencing the use of PSAV and CS (5.55%). Gender, age, income, education, job

Figure 13 shows that companion (4.58%), privacy concerns (4.07%), safety perceptions (3.84%), and cybersecurity awareness (3.81%), all of which highlight users' concerns regarding personal security and data protection in shared mobility services. While the reliability factor demonstrates almost the least important factor in the model.

These findings emphasize that both trip-related factors (e.g., group 1) and personal attributes (e.g., group 2), and safety-related factors (e.g., group 3) strongly influence mode choice. Notably, safety factors as demonstrated in group 3, emerge as significant determinants, reflecting users' hesitations regarding data protection in PSAV services.
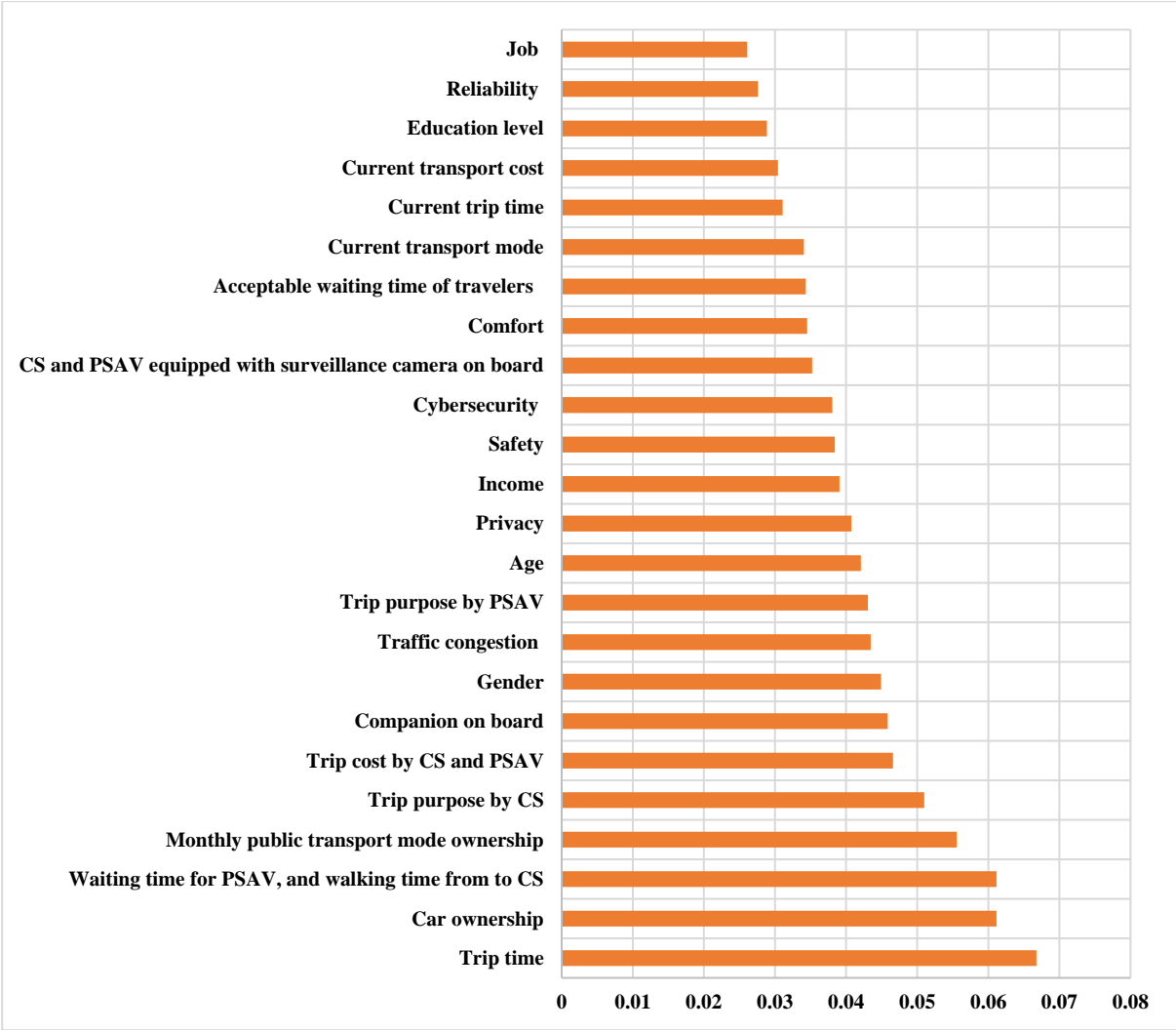
Figure 13. Feature importance of XGBoost.

From a traffic safety perspective, concerns regarding privacy, cybersecurity, and the presence of surveillance cameras emerged as notable determinants in PSAV adoption. These findings align with previous research suggesting that trust in autonomous vehicle systems is a critical factor in adoption decisions [75,76]. Additionally, the strong influence of trip time, cost, and congestion highlights the importance of service efficiency in shaping user choices. In summary, the study underscores the need for policymakers and mobility providers to address safety concerns, optimize travel efficiency, and enhance user trust in PSAV services. Further research could explore the role of real-world safety incidents and regulatory measures in shaping public perception and adoption trends.

4.4. Limitations and Recommendations

While shared autonomous mobility offers promising benefits in terms of congestion reduction and environmental efficiency, ensuring public trust through robust safety regulations and transparency in data security remains imperative. Additionally, the study underscores the need for efficient and cost-effective mobility solutions to enhance the attractiveness of shared transportation options. The insights gained from this study offer valuable implications for policymakers, urban planners, and mobility service providers aiming to optimize shared transportation systems. Future research should explore the long-term impact of real-world PSAV deployment, integrating empirical safety data and user feedback to refine predictive models and enhance public acceptance of autonomous mobility solutions.

5. Conclusions

This study applied machine learning techniques to analyze the behavioral and demographic factors influencing user mode choice between CS and PSAV services, with a particular focus on traffic safety perceptions. The findings provide critical insights into the determinants of mobility preferences, highlighting the interplay between trip characteristics, personal attributes, and safety concerns in shaping user decisions. Among the models evaluated, gradient boosting algorithms, particularly XGBoost, demonstrated superior predictive performance, effectively capturing the complexity of mode choice behavior. The analysis revealed that trip duration, car ownership, walking and waiting times, and trip costs are key determinants of users' mobility preferences. Furthermore, traffic safety concerns, privacy considerations, and cybersecurity risks emerged as significant factors influencing the adoption of PSAVs, underscoring the role of perceived security in shaping attitudes toward autonomous mobility. From a traffic safety perspective, the findings highlight the importance of addressing user apprehensions regarding data privacy, surveillance, and technological reliability in PSAV adoption.

References

1. Sperling, D. Three revolutions: Steering automated, shared, and electric vehicles to a better future; Island Press: 2018.
2. Hancock, P.A.; Nourbakhsh, I.; Stewart, J.J.P.o.t.N.A.o.S. On the future of transportation in an era of automated and autonomous vehicles. **2019**, *116*, 7684-7691.
3. Campisi, T.; Severino, A.; Al-Rashid, M.A.; Pau, G.J.I. The development of the smart cities in the connected and autonomous vehicles (CAVs) era: From mobility patterns to scaling in cities. **2021**, *6*, 100.
4. Axsen, J.; Sovacool, B.K.J.T.R.P.D.T.; Environment. The roles of users in electric, shared and automated mobility transitions. **2019**, *71*, 1-21.
5. Hamadneh, J.; Esztergár-Kiss, D. The preferences of transport mode of certain travelers in the age of autonomous vehicle. *Journal of Urban Mobility* **2023**, *3*, 100054, doi:10.1016/j.urbmob.2023.100054.
6. Jabbie, M.J.I., Innovation; Infrastructure. Rethinking technology sharing for sustainable growth and development in developing countries. **2021**, 936-945.
7. Luca, O.; Andrei, L.; Iacoboaea, C.; Gaman, F.J.S. Unveiling the hidden effects of automated vehicles on "do no significant harm" components. **2023**, *15*, 11265.
8. Chen, Y.; Li, C.; Wang, W.; Zhang, Y.; Chen, X.M.; Gao, Z.J.n.S.M.; Transport. The landscape, trends, challenges, and opportunities of sustainable mobility and transport. **2025**, *2*, 8.
9. Garus, A.; Mourtzouchou, A.; Suarez, J.; Fontaras, G.; Ciuffo, B.J.S.C. Exploring Sustainable Urban Transportation: Insights from Shared Mobility Services and Their Environmental Impact. **2024**, *7*, 1199-1220.
10. Hou, N.; Shollock, B.; Petzoldt, T.; M'Hallah, R.J.I.J.o.S.T. Qualitative insights into travel behavior change from using private cars to shared cars. **2025**, 1-15.
11. Roblek, V.; Meško, M.; Podbregar, I.J.S. Impact of car sharing on urban sustainability. **2021**, *13*, 905.
12. Liyanage, S.; Dia, H.; Abduljabbar, R.; Bagloee, S.A.J.S. Flexible mobility on-demand: An environmental scan. **2019**, *11*, 1262.
13. Hasan, U.; Whyte, A.; Al Jassmi, H.J.A.S.I. A review of the transformation of road transport systems: are we ready for the next step in artificially intelligent sustainable transport? **2019**, *3*, 1.

14. Sharma, A.; Zheng, Z.J.A.C.D., Construction, Operation; Impact, F. Connected and automated vehicles: Opportunities and challenges for transportation systems, smart cities, and societies. **2021**, 273-296.

15. Bathla, G.; Bhadane, K.; Singh, R.K.; Kumar, R.; Aluvalu, R.; Krishnamurthi, R.; Kumar, A.; Thakur, R.; Basheer, S.J.M.I.S. Autonomous vehicles and intelligent automation: Applications, challenges, and opportunities. **2022**, *2022*, 7632892.

16. Chougule, A.; Chamola, V.; Sam, A.; Yu, F.R.; Sikdar, B.J.I.O.J.o.V.T. A comprehensive review on limitations of autonomous driving and its impact on accidents and collisions. **2023**, *5*, 142-161, doi:10.1109/OJVT.2023.3335180.

17. Al Mansoori, S.; Al-Emran, M.; Shaalan, K.J.I.J.o.H.C.I. Factors affecting autonomous vehicles adoption: A systematic review, proposed framework, and future roadmap. **2024**, *40*, 8397-8418.

18. Matin, A.; Dia, H.J.J.o.I.; Vehicles, C. Public perception of connected and automated vehicles: Benefits, concerns, and barriers from an Australian perspective. **2024**.

19. Khan, S.K.; Shiwakoti, N.; Stasinopoulos, P.; Warren, M.J.T.R.P. Cybersecurity Regulations for Automated Vehicles: A Conceptual Model Demonstrating the" Tragedy of the Commons". **2025**, *82*, 3729-3751.

20. Chen, Y.; Khan, S.K.; Shiwakoti, N.; Stasinopoulos, P.; Aghabayk, K.J.R.i.T.B.; Management. Integrating perceived safety and socio-demographic factors in UTAUT model to explore Australians' intention to use fully automated vehicles. **2024**, *56*, 101147.

21. Chen, Y.; Shiwakoti, N.; Stasinopoulos, P.; Khan, S.K.J.S. State-of-the-art of factors affecting the adoption of automated vehicles. **2022**, *14*, 6697.

22. Bala, H.; Anowar, S.; Chng, S.; Cheah, L.J.T.R. Review of studies on public acceptability and acceptance of shared autonomous mobility services: Past, present and future. **2023**, *43*, 970-996.

23. Patel, R.K.; Etminani-Ghasrodashti, R.; Kermanshachi, S.; Rosenberger, J.M.; Foss, A.J.I.J.o.T.S.; Technology. Exploring willingness to use shared autonomous vehicles. **2023**, *12*, 765-778.

24. Jiang, X.; Yu, W.; Li, W.; Guo, J.; Chen, X.; Guo, H.; Wang, W.; Chen, T.J.S. Factors affecting the acceptance and willingness-to-pay of end-users: A survey analysis on automated vehicles. **2021**, *13*, 13272.

25. Ullah, I.; Zheng, J.; Ullah, S.; Bhattarai, K.; Almujibah, H.; Alawad, H.J.S. Unraveling the Complex Barriers to and Policies for Shared Autonomous Vehicles: A Strategic Analysis for Sustainable Urban Mobility. **2024**, *12*, 558.

26. Hamiditehrani, S.; Scott, D.M.; Sweet, M.N.J.T.B.; Society. Shared versus pooled automated vehicles: Understanding behavioral intentions towards adopting on-demand automated vehicles. **2024**, *36*, 100774.

27. Lee, S.; Wang, L.J.T.o.T.S. Consumer Preferences and Determinants of Transportation Mode Choice Behaviors in the Era of Autonomous Vehicles. **2024**, *15*, 37-47.

28. Sadeghpour, M.; Beyazıt, E.J.T.P.; Technology. The new frontier of urban mobility: a scenario-based analysis of autonomous vehicles adoption in urban transportation. **2025**, *48*, 43-66.

29. Sumitkumar, R.; Al-Sumaiti, A.S.J.R.; Reviews, S.E. Shared autonomous electric vehicle: Towards social economy of energy and mobility from power-transportation nexus perspective. **2024**, *197*, 114381.

30. Püschel, J.; Barthelmes, L.; Kagerbauer, M.; Vortisch, P.J.T.R.R. Comparison of discrete choice and machine learning models for simultaneous modeling of mobility tool ownership in agent-based travel demand models. **2024**, *2678*, 376-390.

31. Hu, B.; Tang, J.; Tong, D.; Zhao, H.J.T.B.; Society. Revealing spatiotemporal characteristics of EV car-sharing systems: A case study in Shanghai, China. **2024**, *36*, 100808.

32. Ghorbani, A.; Nassir, N.; Lavieri, P.S.; Beeramoole, P.B.; Paz, A.J.T. Enhanced utility estimation algorithm for discrete choice models in travel demand forecasting. **2025**, 1-28.

33. Yu, D.; Li, Z.; Zhong, Q.; Ai, Y.; Chen, W.J.J.o.A.T. Demand Management of Station-Based Car Sharing System Based on Deep Learning Forecasting. **2020**, *2020*, 8935857.

34. Rahman, M.M.; Paul, K.C.; Hossain, M.A.; Ali, G.M.N.; Rahman, M.S.; Thill, J.-C.J.I.A. Machine learning on the COVID-19 pandemic, human mobility and air quality: A review. **2021**, *9*, 72420-72450.

35. Wu, S.; Falk, K.; Myklebust, T.J.W.E.V.J. Beyond Safety: Barriers to Shared Autonomous Vehicle Utilization in the Post-Adoption Phase—Evidence from Norway. **2025**, *16*, 133.

36. Greenblatt, J.B.; Shaheen, S. Automated vehicles, on-demand mobility, and environmental impacts. *Current Sustainable/Renewable Energy Reports volume* **2015**, *2*, 74-81, doi:10.1007/s40518-015-0038-5.

37. Hao, M.; Yamamoto, T. Shared autonomous vehicles: A review considering car sharing and autonomous vehicles. *Asian Transport Studies* **2018**, *5*, 47-63, doi:10.11175/eastsats.5.47.

38. Hamadneh, J.; Hamdan, N.; Mahdi, A. Users' Transport Mode Choices in the Autonomous Vehicle Age in Urban Areas. *Journal of Transportation Engineering, Part A: Systems.* **2024**, *150*, 04023128, doi:10.1061/JTEPBS.TEENG-7961.

39. Adnan, N.J.T.R.I.P. Exploring the future: A meta-analysis of autonomous vehicle adoption and its impact on urban life and the healthcare sector. **2024**, *26*, 101110.

40. Lee, D.; Hess, D.J.J.H.; communications, s.s. Public concerns and connected and automated vehicles: safety, privacy, and data security. **2022**, *9*, 1-13.

41. Naiseh, M.; Clark, J.; Akarsu, T.; Hanoch, Y.; Brito, M.; Wald, M.; Webster, T.; Shukla, P.J.A.; SOCIETY. Trust, risk perception, and intention to use autonomous vehicles: an interdisciplinary bibliometric review. **2024**, 1-21.

42. Kyriakidis, M.; Sodnik, J.; Stojmenova, K.; Elvarsson, A.B.; Pronello, C.; Thomopoulos, N.J.S. The role of human operators in safety perception of av deployment—insights from a large european survey. **2020**, *12*, 9166.

43. Stoiber, T.; Schubert, I.; Hoerler, R.; Burger, P.J.T.R.P.D.T.; Environment. Will consumers prefer shared and pooled-use autonomous vehicles? A stated choice experiment with Swiss households. **2019**, *71*, 265-282.

44. Zhou, F.; Zheng, Z.; Whitehead, J.; Washington, S.; Perrons, R.K.; Page, L.J.T.R.P.A.P.; Practice. Preference heterogeneity in mode choice for car-sharing and shared automated vehicles. **2020**, *132*, 633-650.

45. Kolarova, V.; Steck, F.; Bahamonde-Birke, F.J. Assessing the effect of autonomous driving on value of travel time savings: A comparison between current and future preferences. *Transportation Research Part A: Policy Practice* **2019**, *129*, 155-169, doi:10.1016/j.tra.2019.08.011.

46. Hamadneh, J.; Esztergár-Kiss, D. The Effects of Multitasking and Tools Carried by Travelers Onboard on the Perceived Trip Time. *Journal of Advanced Transportation* **2021**, *2021*, 5597694, doi:10.1155/2021/5597694.

47. Müller, J.M. Comparing technology acceptance for autonomous vehicles, battery electric vehicles, and car sharing—A study across Europe, China, and North America. *Sustainability* **2019**, *11*, 4333, doi:10.3390/su11164333.

48. Schoettle, B.; Sivak, M. A survey of public opinion about connected vehicles in the US, the UK, and Australia. In Proceedings of the 2014 International Conference on Connected Vehicles and Expo (ICCVE), 2014; pp. 687-692.

49. Howard, D.; Dai, D. Public perceptions of self-driving cars: The case of Berkeley, California. In Proceedings of the Transportation research board 93rd annual meeting, 2014; pp. 1-16.

50. Megens, I.I.; Schaefer, W.W.; van der Waerden, P.P.; Masselink, P.P. Vehicle Users' Preferences Concerning Automated Driving Implications for transportation and market planning. Eindhoven University of Technology, 2015.

51. Payre, W.; Cestac, J.; Delhomme, P. Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation research part F: Traffic Psychology and Behaviour* **2014**, *27*, 252-263.

52. Yan, X.; Levine, J.; Zhao, X. Integrating ridesourcing services with public transit: An evaluation of traveler responses combining revealed and stated preference data. *Transportation Research Part C: Emerging Technologies* **2019**, *105*, 683-696, doi:10.1016/j.trc.2018.07.029.

53. Das, S.; Sekar, A.; Chen, R.; Kim, H.C.; Wallington, T.J.; Williams, E. Impacts of Autonomous Vehicles on Consumers Time-Use Patterns. *Challenges* **2017**, *8*, 32, doi:10.3390/challe8020032.

54. Litman, T. Autonomous vehicle implementation predictions: Implications for transport planning. *Transportation Research Board* **2020**.

55. Lempert, R.; Zhao, J.; Dowlatabadi, H. Convenience, savings, or lifestyle? Distinct motivations and travel patterns of one-way and two-way carsharing members in Vancouver, Canada. *Transportation Research Part D: Transport and Environment* **2019**, *71*, 141-152, doi:10.1016/j.trd.2018.12.010.

56. Fornahl, D.; Hülsmann, M. Markets and Policy Measures in the Evolution of Electric Mobility; Springer: 2016.

57. Efthymiou, D.; Chaniotakis, E.; Antoniou, C. Factors affecting the adoption of vehicle sharing systems. In *Demand for Emerging Transportation Systems*; Elsevier: 2020; pp. 189-209.

58. Shaheen, S.A.; Schwartz, A.; Wipyewski, K. Policy considerations for carsharing and station cars: Monitoring growth, trends, and overall impacts. *Tansportation Research Record* **2004**, *1887*, 128-136, doi:10.3141/1887-15.

59. Ciari, F.; Balac, M.; Balmer, M. Modelling the effect of different pricing schemes on free-floating carsharing travel demand: a test case for Zurich, Switzerland. *Transportation* **2015**, *42*, 413-433, doi:10.1007/s11116-015-9608-z.

60. Perboli, G.; Caroleo, B.; Musso, S. Car-sharing: Current and potential members behavior analysis after the introduction of the service. In Proceedings of the 2017 IEEE 41st annual computer software and applications conference (COMPSAC), 2017; pp. 771-776.

61. Lee, S. Transportation Mode Choice Behavior in the Era of Autonomous Vehicles: The Application of Discrete Choice Modeling and Machine Learning. Portland State University, 2022.

62. Pineda-Jaramillo, J.; Arbeláez-Arenas, Ó.J.J.o.U.P.; Development. Assessing the performance of gradient-boosting models for predicting the travel mode choice using household survey data. **2022**, *148*, 04022007.

63. Shaheen, S.; Bouzaghrane, M.A. Mobility and Energy Impacts of Shared Automated Vehicles: a Review of Recent Literature. *Current Sustainable/Renewable Energy Reports* **2019**, *6*, 193-200, doi:10.1007/s40518-019-00135-2.

64. Fagnant, D.J.; Kockelman, K.M. The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. *Transportation Research Part C: Emerging Technologies* **2014**, *40*, 1-13, doi:10.1016/j.trc.2013.12.001.

65. Greenblatt, J.B.; Shaheen, S. Automated vehicles, on-demand mobility, and environmental impacts. *Current sustainable/renewable energy reports* **2015**, *2*, 74-81, doi:10.1007/s40518-015-0038-5.

66. Elreedy, D.; Atiya, A.F.J.I.S. A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. **2019**, *505*, 32-64.

67. Chen, H.; Cheng, Y.J.I.T.o.I.T.S. Travel mode choice prediction using imbalanced machine learning. **2023**, *24*, 3795-3808.

68. Yang, L.; Shami, A.J.N. On hyperparameter optimization of machine learning algorithms: Theory and practice. **2020**, *415*, 295-316.

69. Alibrahim, H.; Ludwig, S.A. Hyperparameter optimization: Comparing genetic algorithm against grid search and bayesian optimization. In Proceedings of the 2021 IEEE congress on evolutionary computation (CEC), 2021; pp. 1551-1559.

70. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016; pp. 785-794.

71. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A.J.A.i.n.i.p.s. CatBoost: unbiased boosting with categorical features. **2018**, *31*.

72. Chang, X.; Wu, J.; Liu, H.; Yan, X.; Sun, H.; Qu, Y.J.T.A.T.S. Travel mode choice: a data fusion model using machine learning methods and evidence from travel diary survey data. **2019**, *15*, 1587-1612.

73. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems* **2017**, *30*.

74. Yacouby, R.; Axman, D. Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. In Proceedings of the Proceedings of the first workshop on evaluation and comparison of NLP systems, 2020; pp. 79-91.