systems

MDPI

*Article*

# Investigating User Acceptance of Autonomous Vehicles in Developing Cities Using Machine Learning: Lessons from Alexandria, Egypt

**Sherif Shokry [1,*], Ahmed Mahmoud Darwish [2,*], Hazem Mohamed Darwish [3], Omar Elsnossy Ibrahim [4], Maged Zagow [5], Marwa Elbany [6] and Usama Elrawy Shahdah [7]**

1 The Center of Road Traffic Safety, Naif Arab University for Security Sciences, Riyadh 11452, Saudi Arabia
2 Department of Transportation Engineering, Faculty of Engineering, Alexandria University, Alexandria 21533, Egypt
3 Civil Engineering Department, Faculty of Engineering, Alexandria University, Alexandria 21533, Egypt; es_hazem.darwish2025@alexu.edu.eg
4 Faculty of Engineering, Port Said University, Port Said 42526, Egypt; omar150503@eng.psu.edu.eg
5 Department of Architecture Engineering, Faculty of Engineering, Tanta University, Tanta 31527, Egypt; mzagow@f-eng.tanta.edu.eg
6 Department of Civil Engineering, Faculty of Engineering, Port Said University, Port Said 42526, Egypt; mr_elbany@eng.psu.edu.eg
7 Public Works Engineering Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt; usama.shahdah@mans.edu.eg
* Correspondence: sshokry@nauss.edu.sa (S.S.); adarwish@alexu.edu.eg (A.M.D.)

**Abstract**

The willingness to adopt Autonomous Vehicles (AVs) represents a crucial advancement from the sustainable mobility perspective. This is progressively continuing in the developed countries. A comparable shift is expected in developing nations; however, empirical studies remain limited, especially in areas where AVs have not yet been deployed. This study investigates the willingness to adopt AVs in a developing city where AVs have not been deployed yet. A comprehensive travel behavior questionnaire was conducted among local commuters in Alexandria, Egypt, to identify the influential variables affecting AV choice. The well-known machine learning classifier, Extreme Gradient Boosting (XGB), was employed to develop a forecasting model, which indicated a notable accuracy. The results indicated that trip cost was the most influential feature. On the other hand, there is a considerable level of mode captivity, since most travelers prefer to remain with their current mode, regardless of the effects of other variables. A significant share of travelers expressed concerns about shifting to AVs due to safety worries associated with the travel behavior of other transportation modes' commuters. The analysis provides nuanced perspectives on the variables promoting modal shift toward the AVs, supporting future policies for smart urban mobility.

**Keywords:** Autonomous Vehicles (AVs); human-driven vehicles (HDVs); Extreme Gradient Boosting (XGB); Machine Learning (ML); transportation choice modelling

Check for updates

Academic Editors: Eugen Rosca, Radovan Madleňák and Florin Valentin Rusca

Received: 21 November 2025
Revised: 11 December 2025
Accepted: 17 December 2025
Published: 31 December 2025

**Copyright:** © 2025 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.

## 1. Introduction

Choice modelling is one of the key components of travel behavior models, since it defines the distribution of trips among transportation modes utilized for travel based on related variables. These variables include personal characteristics, trip characteristics, and mode characteristics [1,2]. Car ownership and driving license, in addition to household

income, are descriptive examples of personal characteristics [3–7]. Examples of trip characteristics comprise trip purpose and departure time. Additionally, the characteristics of a transport facility include both qualitative and quantitative elements, including system availability, speed, and comfort, for instance [8,9]. Moreover, speed and capacity can be taken into account when describing a transport facility's features [10]. Different study areas prioritize specific factors over others according to local conditions identified through the transport surveys [11,12].

Substantial amount of research and development has been conducted in the area of manufacturing cars, either fully or partially automated. It is anticipated that the development of fully AVs will greatly benefit society by reducing energy consumption and emissions, increasing road capacity utilization, and reducing accidents [13,14]. One innovative application of AVs is that they can coexist with HDVs and improve mobility and accessibility. The adoption of AVs represents a key advancement, particularly within developed countries, in the pursuit of sustainable mobility [15,16]. Limited research, however, is still ongoing, especially in developing countries, on travel behavior modelling in the case of AV presence in the future. Mode choice modeling provides the solution to understand the transport behavior in the future using familiar models or ML techniques.

As a promising approach of Artificial Intelligence (AI), Machine Learning (ML) was utilized in modelling, enabling artificial processing and producing predictions or judgments without the need for explicit programming based on the rich dataset [7,17–19]. Through a process known as training or learning, ML systems are integrated to identify patterns from historical data. This training enables the models to produce predictions about new data. Research studies using machine learning techniques typically consist of two main phases. Carefully choosing relevant variables and models for prediction is the first step in the process. To do this, a portion of the data must be set apart for training and validation purposes.

Basically, supervised and unsupervised approaches are the two main types of ML modelling [20]. Supervised learning is considered when the model is trained with labeled data, allowing each data point to be experienced as an associated correct categorization [21]. Unsupervised learning, however, is employed to distinguish patterns in an input dataset when the preset output variables are not available. The algorithm only identifies hidden patterns from input data; output labels are not included. To verify the target classes discovered by unsupervised learning, the process of manually identifying the true labels is necessary. Because unsupervised learning is inherently subjective and lacks a well-defined objective for response prediction, it is frequently utilized in exploratory settings [22]. Supervised learning models support the interpretation of observed system behavior and enable the prediction of system responses under previously unobserved conditions. These models are capable of capturing complex relationships between primary variables by learning from a given set of labeled data samples.

In this research, ML classifiers are developed and used as tools capable of efficiently forecasting the probability that travelers will select their preferred transportation mode among the available modes versus the innovative one (i.e., AV). The methodology is based on investigating the current travel behavior for travelers in Alexandria City, representing a developing city where the AVs have not yet been introduced in the transportation system, and then determining the key variables affecting choice modelling either toward accepting a modal shift to the AVs, should they be introduced in the future, or toward the current mode. The Extreme Gradient Boosting (XGB) classifier [23] is employed in this study, as it can deal with complex datasets, including multinomial choice modelling by identifying the correlations between independent and dependent variables and determining meaningful thresholds underlying mode-choice behavior [2,24–27].

This research aims to evaluate travelers' behavior and their willingness to shift from HDVs to AVs in developing countries where AVs have not yet been introduced. Accordingly, the study seeks to achieve the following objectives:

- Collect data on travel behavior within the study area through a comprehensive Stated Preference (SP) survey.
- Examine users' current mode choices and evaluate their willingness, along with the underlying factors influencing decision to adopt AVs as an alternative innovative transport mode.
- Model travel behavior using an appropriate machine learning (ML) algorithm that captures mode choice dynamics in the study area.
- Identify the key determinants influencing travel behavior within the developed model, thereby offering detailed insights into mode choice decision making.
- Support future policymaking and infrastructure planning in the smart urban mobility development, facilitating a substantial shift toward AVs when they are integrated into the transportation system.

The research study paper structure is organized as follows: following this introduction section, Section 2 comprehensively delivers the literature review outcomes. Then, Section 3 explains the overall research methodology adopted to achieve the research objectives. The study encompasses a detailed description of the case study, the data collection methodology, and the subsequent statistical analysis and interpretation, which are presented in Section 4. In Section 5, the data analysis and the calibration of the ML model are demonstrated. Section 6 discusses the main results and findings. Finally, the main findings, conclusion and future work are provided in Section 7.

## 2. Literature Review

In many developing countries, the current transportation systems experience considerable complexity due to mode-choice captivity, limited availability, low-income levels, and limited accessibility to public transportation [28–32]. Thus, capturing such complex relationships, especially after introducing innovative and smart mobility alternatives, requires a comprehensive investigation of the context-specific conditions through extensive data preprocessing on the travelers' behavior in the current environment and how their elasticity may change in the future after changing the transportation systems.

Transportation systems in developing countries are dynamic, with frequent changes in policies, economic conditions, and urban development. XGB's flexibility allows it to adapt to these changing conditions and update its predictions accordingly [33,34]. In addition, XGB's ability to construct a series of decision trees and focus on correcting previous errors enhances its predictive power. This is particularly useful in developing countries where the accuracy of predictions can significantly impact transportation planning and policy-making.

Machine learning is progressively applied in modeling the choice of transportation mode since it can capture data structure without prior assumptions, unlike traditional logit models, and is less time-consuming [25,35–39]. Well-known ML classifiers include Decision Tree (DT) [21], Neural Networks [40–42], Support Vector Machine (SVM) [35,43], and Random Forest (RF) [44,45]. The Extreme Gradient Boosting (XGB) classifier is one of the most widely used ML classifiers in many applications, including transportation choice modelling. It is an ensemble technique built on the decision-tree method.

It has demonstrated superior performance compared to other ML classifiers [24], providing better model structures with high predictive power and acceptable levels of overfitting, outperforming the traditional Multinomial Logit (MNL) framework [46,47]. A study conducted by Wang and Ross employed the XGB classifier and the traditional

MNL modelling method to predict choice of transportation mode in Delaware Valley, USA, where the XGB classifier produced better modeling and predicting results than the MNL method [24].

The adoption and public perception of AVs are examined in many research studies. A study conducted by Matin and Dia (2024) focused on the public perception of AVs from an Australian perspective [15], using a large-scale attitudinal survey to assess public concerns, benefits, and barriers to Connected Autonomous Vehicles (CAVs) adoption. The study highlighted demographic differences, especially gender-based concern levels, and emphasized policy suggestions. Unlike our study, it focused more on perceived trust, health-related needs, and privacy concerns than on behavior modeling.

Another study conducted by Aramrattana and Fu (2022) explored the behavioral adaptation using a driving simulator [16]. This experimental approach focused on actual driving behavior among AVs vs. HDVs, identifying changes in lane changes, speed, and following distance. It provided insights into the subconscious behavioral responses rather than stated preferences or survey-based attitudes. This methodology captures short-term interaction effects that may improve safety and operation, in contrast to the predictive and attitudinal models.

Tan et al. (2019) examined AV acceptance from a Chinese perspective [48]. A questionnaire survey was employed to investigate travelers' behavior and the factors determining their willingness to shift to AVs in the future. The traditional logistic regression method was used to model travel behavior. The results emphasized that income, license availability, driving behavior, and the convenience of arriving at a public transportation stop are the main contributing factors. Conversely, travelers who reported a significant degree of current unsafe taxis were more willing to shift to the AVs. Haboucha et al. developed a model for long-term AV decision making by distributing an SP questionnaire to 721 North American residents, using a nested logit kernel model that accounted for panel effects [49].

Collectively, these works provide a triangulated perspective—perception, behavior, and prediction—on AV adoption, offering complementary insights useful for policy, infrastructure planning, and future research. However, the current research investigates the readiness of commuters in Alexandria, Egypt, to adopt AVs in the absence of prior exposure to this technology. This study is distinct for its context, Alexandria as a case from the developing world, and for employing a predictive modeling approach. It utilizes a stated preference survey and machine learning (XGB) to model mode choice behavior, identifying key socioeconomic and travel-related factors influencing AV adoption.

Similarities across all studies include the recognition that public perception, sociodemographics, and experience shape AV acceptance. Differences lie in methodological approaches: our study is data-driven and predictive; the Australian study is policy- and perception-focused; and the Swedish study is experimental, while the other two studies are largely data-driven and predictive. Geographically, our study focuses on a developing city, where AVs have not yet been introduced. By contrast, the Australia, Sweden, and USA studies addressed cases from developed areas. The Chinese study addressed the acceptance of AVs in an area where this technology has already been introduced to the system, unlike our study area. Furthermore, the study implemented a traditional logistic regression modeling approach, while our study utilizes the machine learning (XGB) method to model mode choice behavior. Our study also examines willingness to adopt the AVs under different trip cost scenarios.

In Egypt, a single study conducted by El-Bany [50] in Port-Said City addressed the AV acceptance and usage versus the current transportation systems that were examined through employing the traditional Multinomial Logit (MNL) modelling approach. To the best of the authors' knowledge, all other studies (e.g., [19,25,51–55]) in Egypt investigated

user behavior in choosing the preferred mode without focusing on the travelers' willingness to adopt the AVs.

From previous studies, little work has been conducted to explore the factors affecting mode choice in developing countries with particularly low-income and low awareness of AVs, especially in developing cities where AVs have not yet been introduced, such as Alexandria City, Egypt. These factors may be personal characteristics and/or transport systems. Therefore, this study investigates the factors related to the nature of developing communities that may affect mode choice as a foundation for effectively simulating mode choice behavior effectively. Overall, the research goal is to assess travelers' behavior and willingness to alternate HDVs with AVs in a developing country, where AVs have not yet been introduced. On the other hand, the study employed one of the advanced methods (i.e., XGP ML classifier) in modelling the transportation choice modelling, contrasting with the only available study conducted by El-Bany [50] in Port-Said City, which adopted the traditional MNL modelling approach as mentioned before. The methodology is illustrated in the following part.

## 3. Methodology

The advent of AVs presents a transformative opportunity for sustainable mobility, promising to reshape transportation landscapes worldwide. This study explores travelers' behavior and assesses their willingness to switch from Human-Driven Vehicles (HDVs) to AVs in Alexandria, Egypt, a region where AV technology has not yet been introduced. A mixed-method approach is adopted, combining statistical analysis with machine learning methods to investigate key determinants affecting transportation mode choice behavior and to assess the potential adoption of AVs among travelers in the study area.

This section explains the applied methodology, starting with the data collection process through a Stated Preference (SP) questionnaire, followed by the principles of ML modeling using an XGB classifier.

### *3.1. Rationale*

In this study, XGB is used as an ML classifier since it has been applied in many research studies in the literature and has demonstrated superior flexibility in capturing complex relationships, such as transportation choice modeling, providing reliable results with statistically acceptable modelling performance, including prediction accuracy, overfitting control, and overall robustness [24,27,56].

XGB performs an iterative multi-stage decision process based on the decision tree method. The iterative decision trees gain knowledge from all of preceding trees. The structure of XGB is shown in Figure 1, where the predicted output is calculated using Equation (1) [57]:

$$\hat{yi}(t) = \sum_{k=1}^{n} f_k\,(x_i),\quad f_{k\,\in\,F} \tag{1}$$

where

$F$ represents the space of regression trees.
$f_k$ corresponds to a decision tree indicated by $k$.
$f_k\,(x_i)$ is the result of decision tree $k$.
$y_i$ represents the predicted value of $i_{th}$ instance $x_i$.

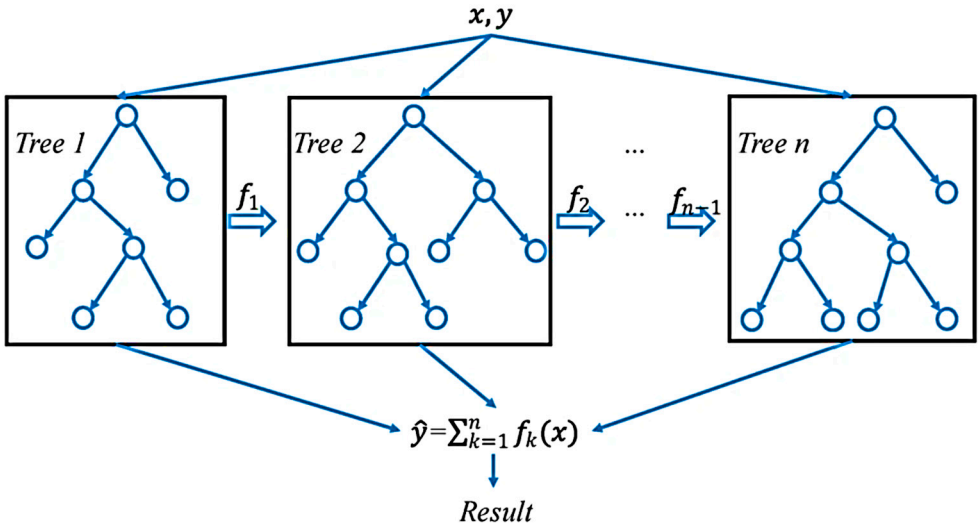



**Figure 1.** The General Structure of the XGB Model [57].

An ensemble model consists of several decision trees combined to produce a model called an XGB Classifier. Each of these decision trees contributes its own portion of improvement on the prediction made by the previous tree by paying attention to samples that had previously been misclassified. In addition, during the splitting process of trees, the algorithm chooses the feature that has the most significant reduction in error for each tree. Ultimately, the total prediction from an XGB model is the result of adding together all of the contributions from all of the decision tree components of the XGB Classifier's Ensemble. By using this type of model, XGB captures nonlinear relationships between features/inputs and does not compromise on its ability to make accurate predictions.

Conventional approaches to choosing modelling, such as the Multinomial Logit (MNL) model, rely on various strong assumptions regarding the development of utility: linearity of utility; Independence from Irrelevant Alternatives (IIA); and additive effects of predictors on utility. All of these assumptions are not very likely to hold for urban travel behavior as it is affected by multiple factors including mode captivity (inability to change modes), income effects, and non-linear cost sensitivity, etc. In contrast to traditional approaches to choosing modelling, XGB does not have any of these structural restrictions and is able to find non-linear relationships and interactions directly from the data. As such, it allows XGB to learn more complex relationships between socio-economic and trip characteristics, and how the interaction of those characteristics affects the way that individuals will respond behaviorally to the introduction of automated vehicles (AVs). Recent studies comparing the predictive accuracy of XGB and MNL models for the same data input have shown that XGB has consistently outperformed MNL in predictive accuracy for mode choice. For example, Wang and Ross (2018) [24] found that XGB was superior to MNL in terms of classification accuracy using real regional data. Therefore, by employing the use of XGB, this study can better account for the complex preference structures that exist within this space, reduce the likelihood of specification bias and provide better forecasts for policy analysis.

The dataset was prepped prior to running the model training, as presented in the research framework in Figure 2, to make sure it would be as accurate and consistent as possible. The responses that had missing information for certain important attributes were removed, as there was only a small portion of missing information (less than 3% of the total sample), and any estimated values would not greatly enhance the coverage area. All of the continuous cost and time values that were extremely outside the norm (outliers) were examined using inter-quartile ranges (IQRs); if they exceeded 1.5 × IQR, we would check against the original data from the survey and accepted those responses only if

they were within the bounds of reasonableness. Categorical variables were represented numerically using one-hot encoding, which allows the XGB algorithm to be able to find and model the interactions (non-linear) that can occur between categories, as opposed to relying on standard numerical encoding, which would not be able to model those non-linear interactions. Continuous variables did not get standardized or normalized since the property tree models do not change regardless of how large or small your variables are and do not need to have their features rescaled. This means that you can use this clean data matrix and recreate any results that have been derived from using the model.

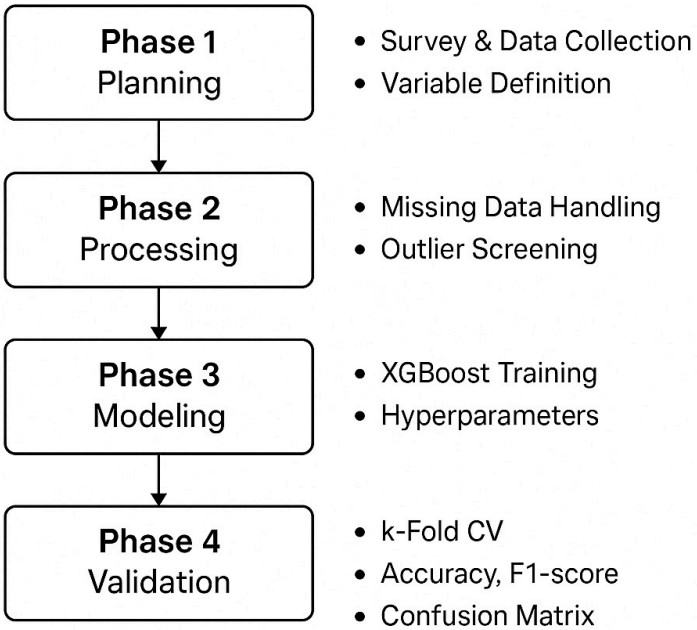



**Figure 2.** Research Framework Flowchart.

### *3.2. Model Validation*

Cross-validation techniques were employed to validate the model and prevent over-fitting. Specifically, the k-fold cross-validation with [k-value] fold method was employed. By splitting the collected sample into $k$ smaller sets, this approach considers training the model on $k - 1$ of those sets and tests it on the remaining part of the data. This process is repeated k times, with each of the k subsets.

This research included using K-Fold Cross Validation of K equal to 10 to verify the models developed through this project. The dataset used for this study was randomly separated into 10 equal parts or "folds" corresponding to the 10 splits of the cross-validation process. Each time the model was trained, 9 of the folds were used to create the training data, while 1-fold was held out as a validation dataset. Each fold was used as a validation dataset for each cycle of training during the entire cross-validation process, at the end of which, an average of the performance metrics collected from the 10-fold sets will be presented as the final results. The process described will allow for a more accurate estimate of the model's capacity for prediction and lessening of the possibility of the model being over-fitted.

### *3.3. Implementation*

The modeling approach is based on implementing the ["XGB" package in R], which provides robust functions for linear models, regularization techniques, and decision tree algorithms. The XGB classifier includes numerous parameters for training the model, which is necessary to build an ML model that can accurately forecast the probability of selecting a

mode by analyzing the fluctuations in various elements within Alexandria City if AVs are added to the list of transportation options.

At the outset, the collected data was subjected to scrutiny and subsequently to a meticulous analysis process including all variables in the questionnaire (i.e., age, gender, income, travel time, etc.). XGB classifiers as an ML tool are used to predict mode choice/willingness to use AVs and HDVs.

The model underwent a series of iterative refinements until a desirable degree of accuracy was attained. After attaining a satisfactory degree of accuracy, the model was employed to predict the proportion of AV use in the future. These forecasts will aid decision-makers in transport development in Alexandria city in the future.

## 4. Case Study and Data Collection

Alexandria was selected as the study area. It is one of the main Egyptian metropolises, with a population of more than 4 million at the beginning of the 21st century. Transportation in Alexandria City is delivered by private and public transportation combining primarily 10 means of transport (i.e., walking, private car, two-wheelers, three-wheelers taxi, and traditional taxis, as private transport means, collective taxis (7 seats and 14 seats), while public transportation includes bus, tram, and train service. The current mobility pattern in Alexandria depends mainly on the use of unsustainable modes of transport, namely collective taxi, in addition to car dependency [51]. This results in daily congestion, where travelers in Alexandria struggle due to the mobility challenges arising from the unsustainable transportation system. On the other hand, there is a lack of awareness about smart transportation systems (e.g., Autonomous Vehicles), as they have not yet been deployed. Because of that, an awareness sheet was added to the questionnaire to introduce AVs and explain their meaning and attributes, ensuring the respondents were aware of the new mode.

There are two types of data to study travel behavior: Revealed Preference (RP) and Stated Preference (SP) interview surveys. The RP interview survey helps understand in understanding the current travel demand; however, it cannot be used directly to assess the future demand under innovative transportation situations that do not yet exist (e.g., the introduction of a new transportation mode to the transportation network). In this regard, the SP method is a suitable approach for implementation in assessing the demand for transportation under hypothetical situations.

### *4.1. Design of the Survey SP Questionnaire*

In order to identify the travelers' behavioral response in hypothetical, non-revealed situations, the Stated Preference (SP) method is employed to gather the data that will form the travel behavior modelling. The objective of this section is to compile information on passenger preferences and demand for both AVs and conventional vehicles. Using effective utility factors and respondents' preferences (i.e., personal data, trip characteristics data, and questions about AVs), the gathered data were used to represent the travelers' behavior in choosing their preferred mode of transport. SP questionnaire was manually designed as an online questionnaire launched between October 2023 to December 2023.

The SP survey form was structured with different types of questions (refer to Table 1):

- Personal Data:
  - Age and gender
  - Driving license availability
  - Education
  - Car ownership and average monthly income
- RP questions about a recently selected trip:

- Trip type and purpose
- Trip start and end.
- Main transportation mode
- Access transportation mode
- Egress transportation mode
- Travel time and travel cost for the main transportation mode
- Travel time and travel cost for the access transportation mode
- Travel time and travel cost for the egress transportation mode
- Waiting time and parking cost

➢ SP questions about the AVs:

- Individuals' concerns about shifting to AVs
- Willingness to shift to AVs, including four scenarios by varying travel cost.

**Table 1.** Descriptive data with ranges and percentages.

| Variables | Ranges | Percent (%) |
|---|---|---|
| ***Personal Data*** | | |
| ➢ *Age* | ➢ <18<br>➢ 18–35<br>➢ 36–65<br>➢ >65 | 4%<br>65%<br>14%<br>17% |
| ➢ *Gender* | ➢ Male<br>➢ Female | 74%<br>26% |
| ➢ *Driving license* | ➢ Yes<br>➢ No | 76%<br>24% |
| ➢ *Education* | ➢ University graduate<br>➢ Technical education graduate<br>➢ Undergraduate student—high education<br>➢ Undergraduate student—technical education<br>➢ High school graduate<br>➢ High school student<br>➢ Elementary stage student<br>➢ Illiterate person<br>➢ Technical education graduate | 23%<br>8%<br>55%<br>0%<br>0%<br>2%<br>1%<br>2%<br>8% |
| ➢ *Car ownership* | ➢ 0 car<br>➢ 1 car<br>➢ 2 cars or more | 22.1%<br>53.6%<br>24.3% |
| ➢ *Income* | ➢ Low<br>➢ Medium<br>➢ High | 16%<br>74%<br>10% |
| ***Current Trip Data*** | | |
| ➢ *Trip type* | ➢ From home<br>➢ To home<br>➢ Other | 55%<br>21%<br>16% |

**Table 1.** *Cont.*

| Variables | Ranges | Percent (%) |
|---|---|---|
| ➢ ***Trip purpose*** | ➢ Work<br>➢ Education<br>➢ Shopping<br>➢ Social<br>➢ Other | 23.1%<br>29.7%<br>13.3%<br>10.2%<br>23.6% |
| ➢ ***Trip start*** | From zone 1 to 17 | |
| ➢ ***Trip end*** | From zone 1 to 17 | |
| ➢ ***WT (waiting time)*** | Time in min | |
| ➢ ***TT (travel time) for Main, Access, and Egress Modes*** | Time in min | |
| ➢ ***TC (travel cost) for Main, Access, and Egress Modes*** | Cost in EGP | |
| ➢ ***PC (parking cost)*** | Cost in EGP | |
| ➢ ***Main Transportation Mode*** | ➢ Walking<br>➢ Motorcycle<br>➢ Private Car<br>➢ Taxi<br>➢ Tram<br>➢ Railway<br>➢ Shared Taxi-14 Seats<br>➢ Shared Taxi-7 Seats<br>➢ Bus<br>➢ Three Wheelers Taxi<br>➢ Other | 3.5%<br>0.9%<br>14.7%<br>4.6%<br>4.6%<br>2.3%<br>61.9%<br>0.8%<br>5.1%<br>1.0%<br>0.6% |
| ➢ ***Access Mode*** | ➢ Walking<br>➢ Motorcycle<br>➢ Taxi<br>➢ Shared Taxi-14 Seats<br>➢ Shared Taxi-7 Seats<br>➢ Three Wheelers Taxi<br>➢ Other | 72.1%<br>0.6%<br>2.9%<br>10.3%<br>3.6%<br>6.3%<br>4.2% |
| ➢ ***Egress Mode*** | ➢ Walking<br>➢ Motorcycle<br>➢ Taxi<br>➢ Shared Taxi-14 Seats<br>➢ Shared Taxi-7 Seats<br>➢ Three Wheelers Taxi<br>➢ Other | 72.8%<br>0.8%<br>2.1%<br>11.4%<br>2.4%<br>3.2%<br>7.2% |

**Table 1.** *Cont.*

| Variables | Ranges | | Percent (%) |
|---|---|---|---|
| ***Stated Transportation Mode*** | | | |
| ***Individuals' concerns about shifting to AVs*** | ➢ Not Familiar | | 12.9% |
| | ➢ Increased Cost | | 35.0% |
| | ➢ Electricity charging Issues | | 9.3% |
| | ➢ Driving Behavior of Other Modes | | 30.9% |
| | ➢ Resale Issue | | 2.0% |
| | ➢ Other | | 9.8% |
| ***Willingness to shift to AVs*** | If the trip cost increases by 100% | Yes | 13.9% |
| | | No | 86.1% |
| | If the trip cost increases by 50% | Yes | 29.7% |
| | | No | 70.3% |
| | If the trip cost increases by 20% | Yes | 60.6% |
| | | No | 39.4% |
| | If the trip cost does not increase | Yes | 78.3% |
| | | No | 21.7% |

### 4.2. Sample Size

Discrete-Choice Experiments (DCEs), which address a variety of policy issues, are becoming a widely used tool in mode choice analysis. The required minimum sample size, DCE's suggested general rule of thumb sample size requirement: The most popular guideline for figuring out the necessary sample size for a DCE was put forth by [58]. According to the rule, there should be enough respondents to meet the inequality according to Equation (2):

$$N \geq 500\, L^{max} / JS \quad (2)$$

Let $L^{max}$ denote the maximum number of levels across all attributes, $J$ represent the number of alternatives per choice task (excluding the "None" option), and $S$ is the number of choice sets presented to each respondent. The total number of two-way interaction terms in the utility specification corresponds to the maximum product of the number of levels between any two attributes. Here, $L = 19$, $J = 11$, $S = 1$, so the minimum sample size needed is 792. The collected dataset includes 1045 successful responses. By comparing the collected samples with the minimum sample size, the samples are suitable for choice modelling.

Despite having a population that exceeds 4 million people, the number of respondents that should be drawn from Alexandria's population for use in responding to the Stated Preference DCE is based primarily on the complexity of the DCE design and the number of parameters to be estimated. While the total number of 1045 respondents exceed the minimum number required to meet the design requirements, it is within the range of sizes typically used for AV acceptance and mode-choice studies. Therefore, it can be concluded that, although it will not adequately address every sub-group or segment of Alexandria's population, this sample will provide a statistically viable basis for analysing behavioural data.

### 4.3. Descriptive Analysis of the Collected Data

To simulate travel behavior effectively, it is important to identify factors affecting the choice behavior among AVs and HDVs. As mentioned before, a Stated Preference (SP) survey was designed to determine the current preferences of travelers when choosing the current transportation mode as well as their willingness to choose the AV as an innovative transportation option over the currently selected mode. Table 1 depicts the description of each factor along with its percentage of the sample. The data include continuous variables as well as categorical variables. For example, due to the discrete and multinomial nature of

some independent variables, the choice sets include 11 categories for the main mode and 7 categories for the access mode and egress mode questions, while the travel time and cost are continuous variables.

## 5. Data Analysis & Modeling Process

All factors were used to model the choice. Once the model reaches an acceptable accuracy rate, it is utilized to forecast the willingness to use AVs and HDVs within the city. These choice predictions provide valuable guidance to policymakers in the development of new transportation networks and the promotion of autonomous mobility alternatives. The independent variables are grouped into two main clusters (refer to Table 2): categorical variables and continuous variables. The main descriptive statistics of the collected samples for the categorical variables are as follows:

- 74% of the surveyed individuals were male, while 26% were female.
- The largest proportion of users was from 18 to 35 years old.
- 76% held a driving license.
- 23% of respondents were university graduates, while 55% were undergraduate students.
- Work and educational trips accounted for approximately 53.2% of all trips.
- 22% of respondents did not own a private car, 53.6% owned one car, and the remaining respondents owned more than one car.
- 74% of respondents were from the medium-income level.
- The Shared Taxi-14 Seats was the main transportation mode with a 61% share, followed by the private car a 14% share.
- 55% of trips were home-based.
- 35% of respondents were worried about shifting to AVs due to the increased cost, followed by the driving behavior of other modes (31%).
- 78% of respondents accepted the modal shift to the AVs if the trip cost did not increase, while this share dropped to only 14% if the trip cost doubled.

**Table 2.** Model variables with type.

| | Variable | Type |
|---|---|---|
| **1** | Gender | Categorical |
| **2** | Age | |
| **3** | Driving license availability | |
| **4** | Household car ownership | |
| **5** | Household income | |
| **6** | Education | |
| **7** | Trip type | |
| **8** | Trip purpose | |
| **9** | Revealed main transportation mode | |
| **10** | Main mode trip cost (LE) | Continuous |
| **11** | Main mode trip time (min) | |
| **12** | Access transportation mode | Categorical |
| **13** | Access mode trip cost (LE) | Continuous |
| **14** | Access mode trip time (min) | |
| **15** | Egress transportation mode | Categorical |
| **16** | Egress mode trip cost (LE) | Continuous |
| **17** | Egress mode trip time (min) | |
| **18** | Waiting time (min) | |
| **19** | Parking cost (LE) | |
| **20** | AVs trip cost (LE) | |
| **21** | Reasons for worry about shifting to AVs | Categorical |
| **22** | Future chosen transportation mode after introducing AVs | |

For the continuous variables, histograms of the trip characteristics (trip time and trip cost) for the three primary transportation modes of the individual trip (main mode, access mode, and egress mode) were generated in statistical analysis to illustrate the distribution of variables. Parking cost and waiting time were also included. They were effectively analyzed in the case of continuous variables such as trip time and cost (examples are shown in Figure 3a–e), which indicate, for example, the variation in travel time for the main, access, and egress transportation modes. Table 3 presents the descriptive statistics for continuous variables.

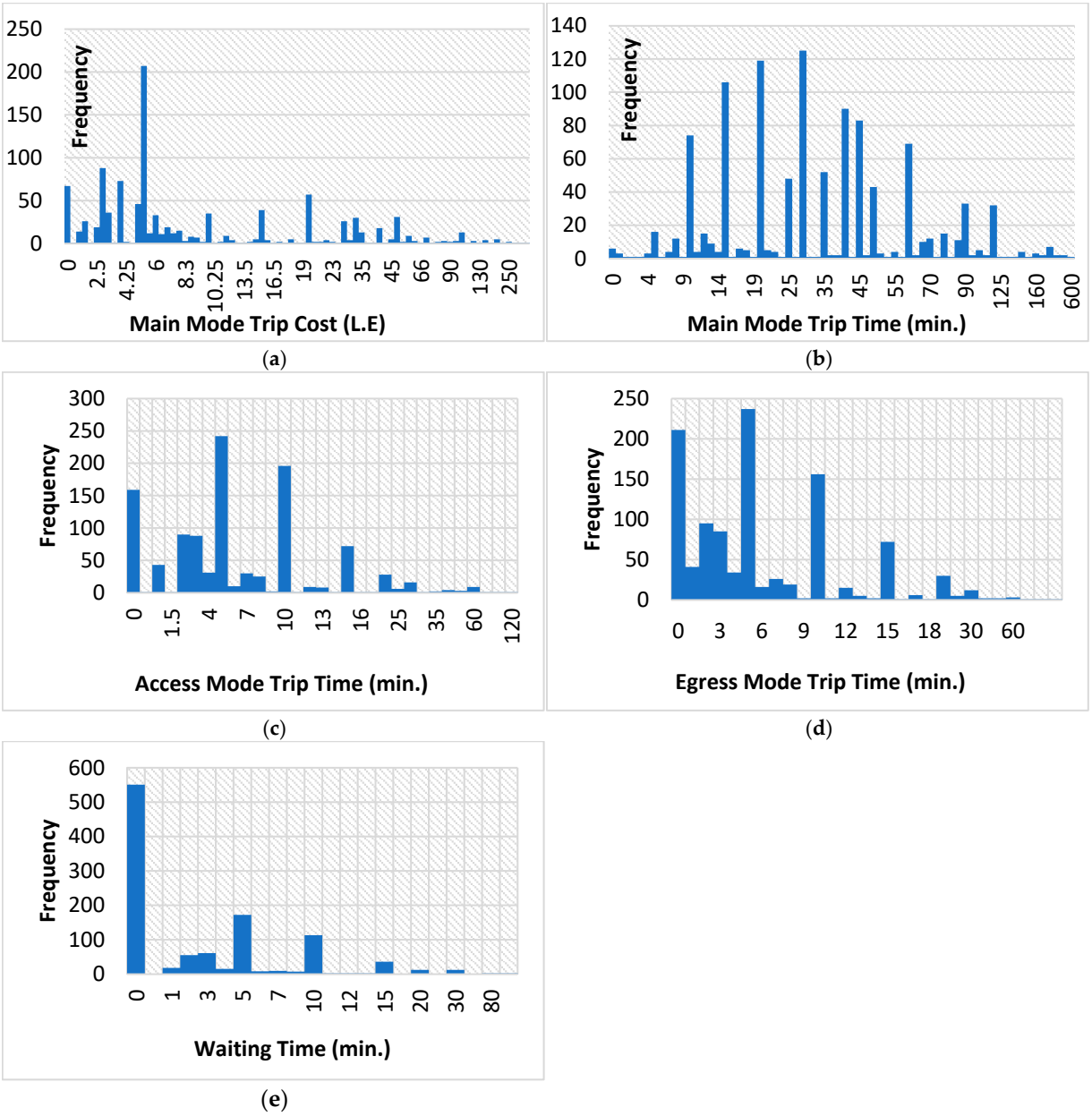



**Figure 3.** (**a**): Main Mode Trip Cost (L.E); (**b**): Main Mode Trip Time (min.); (**c**): Access Mode Trip Time (min.); (**d**): Egress Mode Trip Time (min.); (**e**): Waiting Time (min.).

**Table 3.** Descriptive statistics for continuous variables.

| Continuous Variable | Mean | Std. Dev. | Min | Max | Skewness |
|---|---|---|---|---|---|
| **Main Mode Trip cost (LE)** | 17.85549 | 47.56207 | 0 | 1000 | 12.62794 |
| **Main Mode Trip Time (min.)** | 39.86491 | 36.37561 | 0 | 600 | 4.767436 |
| **Access Mode Trip cost (LE)** | 2.507849 | 12.1174 | 0 | 250 | 12.13988 |
| **Access Mode Trip Time (min)** | 7.421976 | 9.493604 | 0 | 120 | 4.581215 |
| **Egress Mode Trip cost (LE)** | 2.172022 | 13.90268 | 0 | 350 | 17.26126 |
| **Egress Mode Trip Time (min.)** | 6.442699 | 8.055673 | 0 | 120 | 5.224567 |
| **Waiting Time (min.)** | 3.899815 | 8.043427 | 0 | 120 | 7.776572 |
| **Parking cost (LE)** | 0.613158 | 3.160362 | 0 | 60 | 9.777643 |
| **Autonomous Vehicles Trip cost (LE)** | 32.98664 | 81.3165 | 0 | 2000 | 10.42466 |

## 6. Modelling Results and Discussion

The analyses provided insights into the dataset under investigation. The outcomes synthesize the results from the different models, taking into account the performance metrics, feature importance, and model characteristics. Through the employment of XGB analysis, we modeled and interpreted the dynamics of factors affecting the target variable, associated with trip data. The XGB model was implemented using the "XGB" package in R software 4.3.2.

Several performance indicators were calculated and assessed to evaluate the performance of the calibrated model, including prediction capability, recall, precision, F1-score, and confusion matrix [37,39,59–62]. Prediction capability is the degree of accuracy of the predicted samples from the observed ones. t represents the proportion of correctly simulated choices relative to those selected by respondents. The recall represents the percentage of correctly predicted transportation mode choices over the total number of observed choices of the same transportation mode. Likewise, precision is the proportion of total correctly simulated samples of a transportation mode over total simulated choices of the same mode. However, F1 score combines both rates of precision and recall to provide a better representation of the model's performance regarding each transportation mode choice simulation.

### *6.1. Feature Importance for XGB Classifier*

The feature importance values listed indicate how much each feature contributes to the model's predictions, as shown in Figure 4. The most important feature is the trip cost, which confirms that the choice of transportation mode in developing countries is highly correlated with transportation cost, because of relatively low-income levels. This is followed by other features with noticeable importance including the revealed transportation mode, which represents a key driver in selecting a stated mode once AVs are introduced into the network. This finding indicates that, as many travelers in this area in a developing country make their future decisions based on their current decisions, suggesting a high level of mode captivity. That is, regardless of other variables, travelers make their decisions based on their biases toward their current mode. Other features show comparable importance, such as waiting time, reasons for worry about shifting to AVs, income, education, and access mode. These variables highlight the need to increase travelers' awareness of this new mode and its potential benefits.

### *6.2. Accuracy and Confusion Matrix*

When evaluating the model performance, accuracy and the confusion matrix were examined. Accuracy was 88.3% for the training set and 85.2% for the testing set, which implies that the classifier generated correct predictions for all instances in the dataset with acceptable overfitting level. Table 4 shows the precision, recall, F1-score, and support

sample for each of the transportation modes. Most data points were classified accurately, and misclassifications were limited.

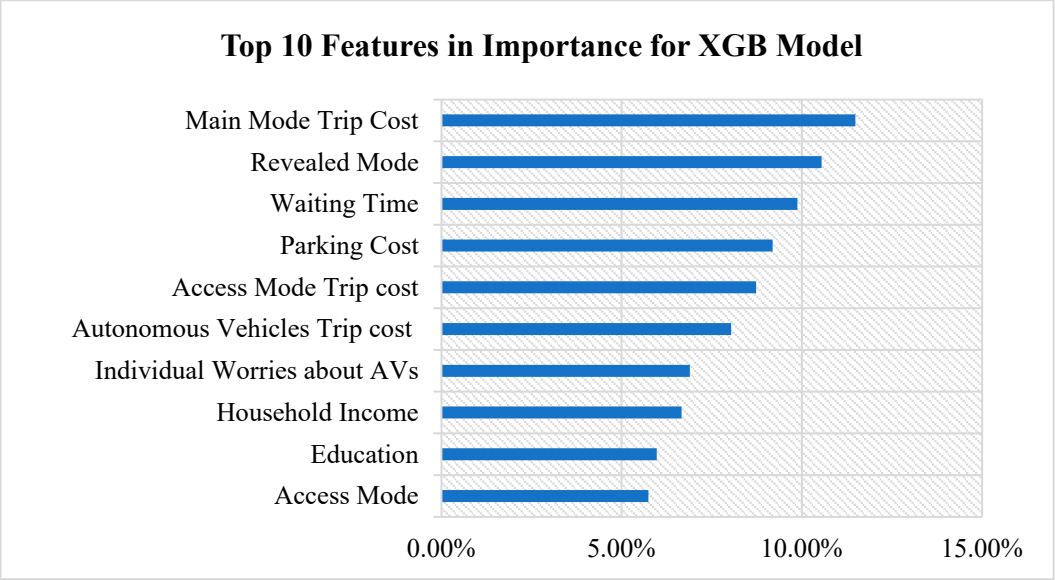



**Figure 4.** Feature importance for XGB classifier.

**Table 4.** Model Results.

| Mode | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| **Autonomous Vehicles** | 0.87 | 0.83 | 0.85 | 476 |
| **Taxi** | 0.74 | 0.77 | 0.75 | 30 |
| **Motorcycle** | 0.50 | 0.50 | 0.50 | 4 |
| **Other** | 0.67 | 0.67 | 0.67 | 3 |
| **Private Car** | 0.83 | 0.77 | 0.80 | 82 |
| **Railway** | 0.92 | 0.86 | 0.89 | 14 |
| **Shared Taxi-14 Seats** | 0.81 | 0.87 | 0.84 | 352 |
| **Shared Taxi-7 Seats** | 0.71 | 1.00 | 0.83 | 5 |
| **Bus** | 0.78 | 0.83 | 0.81 | 30 |
| **Three Wheelers Taxi** | 0.90 | 1.00 | 0.95 | 9 |
| **Tram** | 0.85 | 0.73 | 0.79 | 30 |
| **Walking** | 0.67 | 0.60 | 0.63 | 10 |

A confusion matrix is a table that summarizes a classification algorithm's results in summary form [63]. In the proposed model, as shown in Table 5, the confusion matrix is a ($12 \times 12$) matrix, where the rows represent the actual class labels, and the columns represent the predicted class labels. The elements on the diagonal indicate the correct classifications, with an average accuracy rate of approximately 85%, as mentioned before. Since there are many zeros in the off-diagonal elements, this further confirms that there were limited misclassifications. The classifier achieved strong accuracy. In summary, the confusion matrix confirms that the model provides a highly successful classification performance.

### *6.3. Model Overfitting*

When examining the training and validation accuracies, the provided values suggest satisfactory error levels for both the training and validation sets. The training accuracy is a very high value, which indicates a strong fit of the model to the training data. The model's predictions closely match the actual values in the training set. Similarly, for the validation set, the validation accuracy indicates that the model's performance is also acceptable on unseen data.

**Table 5.** Confusion Matrix.

| Transportation Mode | Autonomous Vehicles | Taxi | Motorcycle | Other | Private Car | Railway | Shared Taxi-14 Seats | Shared Taxi-7 Seats | Bus | Three Wheelers Taxi | Tram | Walking | Modeled | Observed |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Autonomous Vehicles** | 409 | 1 | 8 | 4 | 2 | 1 | 25 | 2 | 7 | 1 | 0 | 3 | 409 | 476 |
| **Taxi** | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3 |
| **Motorcycle** | 7 | 0 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 23 | 30 |
| **Other** | 8 | 0 | 0 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 22 | 30 |
| **Private Car** | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 4 |
| **Railway** | 2 | 0 | 0 | 0 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 14 |
| **Shared Taxi-14 Seats** | 39 | 0 | 0 | 0 | 0 | 0 | 313 | 0 | 0 | 0 | 0 | 0 | 313 | 352 |
| **Shared Taxi-7 Seats** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 5 | 5 |
| **Bus** | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 25 | 0 | 0 | 0 | 25 | 30 |
| **Three Wheelers Taxi** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 0 | 0 | 9 | 9 |
| **Tram** | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 63 | 0 | 63 | 82 |
| **Walking** | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 6 | 10 |

Overfitting, on the other hand, happens when a model fits the training set too closely, capturing anomalies and noise that don't transfer to new data. The fact that the training and validation errors are both low and similar indicates that the model is not overly complex and can make accurate predictions on both the data it was trained on and new data [64]. Thus, the developed model does not show signs of overfitting and performs consistently on the validation set.

### *6.4. Results, Interpretation and Discussion*

The results of the machine learning analysis and modeling provide a nuanced understanding of the (AV) determinants influencing transportation mode choices in developing urban environments such as Egypt. These insights are crucial for urban planning and transport policy-making, enabling the development of strategies that are data-driven and responsive to the needs of the population. According to the results, sustainable transportation and AV-oriented policies can be implemented for Alexandria, Egypt. This can be summarized in the following points:

A significant share of travelers expressed concerns about shifting to AVs due to concerns about the behavior of other road users. This requires improvements in the driving behavior standards in developing countries before introducing AVs. Driving regulations should be applied strictly. This also highlights the importance of non-motorized travel, such as walking or biking to work or educational institutions. This would necessitate policies that create safe and efficient routes for pedestrians and cyclists to reduce the ever-increasing demand for HDVs. Infrastructure improvements, such as dedicated bike lanes and separated public transportation routes, should be targeted around educational institutions and employment hubs.

The modeling results indicate that there is a considerable level of mode captivity in the study area. Most travelers make future mode-choice decisions by refusing to change their current mode regardless of the effects of other variables. Thus, understanding the reasons behind such captivity is vital in developing a meaningful modal shift to a new innovative mode.

Education: The modeling results indicate that the willingness to embrace AVs over traditional HDVs increases as education level rises. Funding and financial support for awareness campaigns about the benefits of AVs can encourage a shift toward to this mode of transport.

Household income: As income level increases, the ability to afford a shift to AVs rises. This is also reflected in the model results for the main-mode trip-cost variable. Logically, households that currently afford higher trip costs are more willing and able to shift to AVs. Policies can support this by supporting the wide production and market expansion of AVs along with subsidies for AV systems. This would help bridge the affordability gap for AVs in Alexandria City.

AV travel cost: The model shows that willingness to shift to AVs decreases as monetary costs increase, including in-vehicle and parking cost. From a policy perspective, it is better to reduce parking hassles to make AVs more attractive. This could make AVs a more convenient option for running errands or going to appointments in busy areas. Intelligent traffic management, priority policies at intersections, and mobile Apps with live tracking may improve the innovative mode attractiveness.

Access-mode trip cost is influenced by the travel distance required to reach the main mode. As this generalized cost increases, willingness to shift to AVs declines. Planners should develop first and last-mile solutions to bridge the gap between home and transit stations, potentially integrating AVs as a service in these segments. Empowering feeder systems and last-mile trips, which reduce access-mode trip cost, will improve on the feasibility of using the AV system.

The modeling results show a lack user experience with the perceived benefits of AVs, where more than 30% of travelers were unfamiliar with them. Policymakers should design incentives to encourage the use of AVs for specific trip types, where a high likelihood of AV preference can be identified. Education and awareness campaigns should be tailored to address this unfamiliarity with AVs, thereby increasing public acceptance and usage.

Trip type and destination: The significance of the 'Trip Type to/from Home' variable suggests that most travel is centered on commuting to and from home. The model highlights the importance of home-based trips. This insight can guide better planning of residential areas, ensuring they are well-connected to various modes of transport, which may enhance the attractiveness of AVs for home-based trips.

Education/work-related travel: These are the primary trip purposes. This may prompt the city to enhance transportation services around educational facilities (i.e., schools, universities, and other educational institutions) as well as major employment centers.

Mode integration: The prediction values at the leaf nodes of the decision tree can guide the integration of AVs into the current transport network, ensuring that new technologies complement the existing HDVs.

Although several machine-learning methods are available for mode choices such as Random Forest, SVM, and Neural Networks—this study did not aim to compare algorithms. The focus was to identify the key determinants of behaviour under hypothetical AV scenarios. XGB was chosen because it performs well with structured tabular data, offers strong predictive ability, and provides clear feature-importance outputs that support interpretation. Running and tuning multiple models would have added complexity without improving the core findings. A comparison across algorithms is valuable, but it is more appropriate as follow-up work using the same dataset and controlled conditions.

The survey provides several valuable insights beyond purely practical considerations. When the costs of the various modes are equal, acceptance of the AV reaches 80% though acceptance of the AV levels declines quickly as costs of alternatives rise. For Zamzam, this

indicates a positive correlation between acceptance of AVs and a competitive degree of regulatory price control.

Second, the degree of mode captivity revealed in the findings is noteworthy. Among many user segments, an Active Mode is entrenched, and a preference exists to continue using the Active Mode despite additional costs, inconvenience, and the availability of an AV. This suggests that while touchless technical solutions are provided, other efforts will be required to induce travel behavior to shift. We propose soft measures to reduce user inertia, using various devices to integrate user visibility of the AV's benefits.

Third, while the AV technology and operational safety gaps drew the greatest number of survey comments, safety concerns about the road discipline of other drivers and AV folk remain. Building regulatory trust will be necessary to provide road discipline, safety, and AV–HDV Mode separation.

Finally, there are differences in acceptance of AVs by socio-demographic group. Higher socio-economic, and therefore higher level of education, groups—the early adopters of AVs—are more likely to support slow convergence.

In conclusion, AVs acceptance exists and is greatest when there is price, travel, and overall convenience.

## 7. Conclusions

Transportation automation is anticipated to significantly influence various aspects of daily life. While it has already been implemented in many developed countries—where it contributes to improved transportation efficiency and mobility—its potential impact in developing nations remains uncertain. To assess whether individuals would adopt such a novel transportation option based on their specific travel needs and influencing factors, forecasting user choice behavior is essential. Therefore, the primary objective of this study is to model mode choice behavior if AVs are introduced into Alexandria's transportation system.

The proposed mode choice model was developed using many key factors including personal characteristics as well as trips and transportation modes features. To predict mode choice and willingness to adopt AVs versus HDVs, the study employed the Extreme Gradient Boosting (XGB) classifier, a machine learning (ML) technique. The dataset comprised 1045 valid responses. A feature importance analysis was conducted to identify the most influential input variables affecting the mode choice predictions. The results conclude that:

- From the feature importance analysis, there is considerable importance associated with several variables, including main mode trip cost, revealed transportation mode, waiting time, parking cost, access mode trip cost, AV travel cost, individual concerns about AVs, household income, education level, and access mode.
- The model achieved an accuracy of 85.2% for the testing set, which indicates that the classifier made correct predictions for most instances in the dataset.
- The model achieved an average F1-score of 0.83, which is considered acceptable. This indicates that the model does not suffer from overfitting.
- Since there are many zeros in the off-diagonal elements of the confusion matrix, the model shows limited misclassifications. Thus, the XGB classifier achieved strong accuracy. In summary, the confusion matrix confirms that the model delivers a highly successful classification.

By leveraging the predictive power of decision tree analyses, urban planners and policymakers in Egypt can make informed decisions that not only facilitate the smooth integration of AVs into the transportation landscape but also optimize existing HDVs to meet the dynamic needs of urban residents. The goal is to create a cohesive, efficient, and

user-friendly transportation ecosystem that supports the economic and social fabric of the country.

It is important for future analysis to consider the effect of modal captivity and current user experience with AVs together alongside local socio-economic characteristics in developing countries.

If environmental factors are included in the dataset and prove significance, policies could promote sustainable transportation options such as electric buses or carpooling to reduce carbon emissions.

This comprehensive approach will ensure that the transition to AVs contributes positively to the urban fabric, balancing technological possibilities with the unique economic, social, and cultural dynamics of developing countries. This choice-prediction capability can provide valuable guidance to policymakers in the development of sustainable transportation networks and the promotion of autonomous mobility alternatives.

In-depth analysis is recommended to further investigate the current captivity of transportation mode choice and its determinants, whereby additional variables may be added to the calibrated model. In addition, it is recommended to examine different social demographics to assess whether there is any evidence of heterogenous samples, which may support the development of multiple classifying models for the different social groups or clusters.

The study is conducted using an online self-reported stated preference. Therefore, the sample is biased towards a younger and more educated group with access to a car or another private mode of transport, which will affect how the respondents respond to questions regarding their willingness to use AVs. This means that the study may not be able to accurately reflect all segments of society and some of the potential reasons for this include that the study is primarily based on socio-economic and trip-related characteristics; however, there are many other reasons (and desire) that could affect how a person decides which mode of transport to use, and the study does not incorporate those other factors.

In order to address the limitations outlined in the study, we suggest that future studies incorporate the following recommendations:

(a) Use a stratified or quota-based sampling methodology to ensure better representation of the overall population.

(b) Utilize hybrid choice or structural equation models to include attitudinal and psychological variables.

(c) Use various modelling methodologies to validate and compare prediction techniques and results.

(d) Provide longitudinal follow-up studies (and/or integrate stated and revealed preferences) once AVs are available to the public.

Acceptance and modal captive restrictions continue to be significant components of the adoption of upstream-cost-driven pricing for AVs. Therefore, the two policy issues identified based on the empirical findings above are (1) the importance of developing cost-sensitive pricing mechanisms (e.g., introductory discounts, capped fares, bundled mobility packages) to remain competitive with existing options; and (2) the need to mitigate (i.e., replace) behavioural inertia via targeted induction. Examples of targeted induction include pilot projects along with the provision of AV service on selected campuses or employment zones and the provision of financial incentives to those who are willing to be 'early adopters' of AV technology. Finally, given the role of access cost in the decision to adopt AVs, it is advisable that the integration of the first/last mile continues to occur using AV shuttle feeders between transit hubs and high-density residential areas. The above actions represent a definitive 'translation' of the empirical evidence into an actionable way forward for developing an effective implementation plan for AVs.

**Author Contributions:** Conceptualization, A.M.D.; software, O.E.I.; validation, M.Z.; formal analysis, A.M.D. and H.M.D.; investigation, A.M.D.; methodology, A.M.D.; resources, H.M.D.; data curation, A.M.D. and H.M.D.; writing—original draft preparation, A.M.D., H.M.D., O.E.I., M.Z., M.E., S.S. and U.E.S.; writing—review and editing, A.M.D., M.Z., M.E., S.S. and U.E.S.; visualization, A.M.D.; supervision, A.M.D. and S.S.; project administration, S.S.; funding acquisition, S.S.; resource coordination, S.S.; full management of the revision process, S.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data will be provided upon request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| AVs | Autonomous Vehicles |
| HDVs | Human-Driven Vehicles |
| XGB | Extreme Gradient Boosting |
| ML | Machine Learning |

## References

1. Qin, H.; Yu, B.; Zhang, Y. Exploring commuters' mode preference to autonomous vehicles based on a personalized travel experience survey. *Transportation* **2024**, 1–34. [CrossRef]
2. Liu, J.; Xiao, L. Socioeconomic differences in effect size: Predicting commuting mode choice of migrants and locals using a light gradient boosting approach. *Transportation* **2024**, *51*, 1–24. [CrossRef]
3. Tan, L.; Ma, C. Choice behavior of commuters' rail transit mode during the COVID-19 pandemic based on logistic model. *J. Traffic Transp. Eng. (Engl. Ed.)* **2021**, *8*, 186–195. [CrossRef]
4. Zagow, M.; Elbany, M.; Darwish, A.M. Identifying urban, transportation, and socioeconomic characteristics across US zip codes affecting CO2 emissions: A decision tree analysis. *Energy Built Environ.* **2024**, *6*, 484–494. [CrossRef]
5. Abdullah, M.; Ali, N.; Bilal Aslam, A.; Javid, M.A.; Hussain, S.A. Factors affecting the mode choice behavior before and during COVID-19 pandemic in Pakistan. *Int. J. Transp. Sci. Technol.* **2022**, *11*, 174–186. [CrossRef]
6. Shokry, S.; Alrashidi, A.; El-Bany, M.E.-S.; Darwish, A.M. Impact of road geometry on school-area traffic congestion using regression and machine learning analysis: Lessons from six Saudi cities. *Transp. Res. Interdiscip. Perspect.* **2025**, *34*, 101686. [CrossRef]
7. Narayanan, S.; Makarov, N.; Magkos, E.; Grau, J.M.S.; Aifadopoulou, G.; Antoniou, C. Can Bike-Sharing Reduce Car Use in Alexandroupolis? An Exploration through the Comparison of Discrete Choice and Machine Learning Models. *Smart Cities* **2023**, *6*, 1239–1253. [CrossRef]
8. Elkafoury, A.; Zagow, M.; Saeed, K.; Darwish, A.M. Model Willingness to Use Public Transport in the USA Based on Socio-Economic and Demographic Characteristics. *Civ. Eng. Archit.* **2023**, *11*, 1487–1497. [CrossRef]
9. Darwish, A.M.; Zagow, M.; Elkafoury, A. Impact of land use, travel behavior, and socio-economic characteristics on carbon emissions in cool-climate cities, USA. *Environ. Sci. Pollut. Res.* **2023**, *30*, 91108–91124. [CrossRef]
10. Alsobky, A.; Darwish, A.M. A realistic framework for volume-delay function determination. *Ain Shams Eng. J.* **2023**, *14*, 102279. [CrossRef]
11. Ortúzar Jde, D.; Willumsen, L.G. *Modelling Transport*, 4th ed.; Wiley-Blackwell: Chichester, UK, 2011. [CrossRef]
12. Garus, A.; Mourtzouchou, A.; Suarez, J.; Fontaras, G.; Ciuffo, B. Exploring Sustainable Urban Transportation: Insights from Shared Mobility Services and Their Environmental Impact. *Smart Cities* **2024**, *7*, 1199–1220. [CrossRef]
13. Milakis, D.; Van Arem, B.; Van Wee, B. Policy and society related implications of automated driving: A review of literature and directions for future research. *J. Intell. Transp. Syst.* **2017**, *21*, 324–348. [CrossRef]
14. Saravanos, A.; Pissadaki, E.K.; Singh, W.S.; Delfino, D. Gauging Public Acceptance of Conditionally Automated Vehicles in the United States. *Smart Cities* **2024**, *7*, 913–931. [CrossRef]
15. Matin, A.; Dia, H. Public perception of connected and automated vehicles: Benefits, concerns, and barriers from an Australian perspective. *J. Intell. Connect. Veh.* **2024**, *7*, 108–128. [CrossRef]

16. Aramrattana, M.; Fu, J.; Selpi, J. Behavioral adaptation of drivers when driving among automated vehicles. *J. Intell. Connect. Veh.* **2022**, *5*, 309–315. [CrossRef]
17. Li, H.; Jin, Z.; Cui, H.; Tu, H. An exploration of the preferences and mode choice behavior between autonomous demand-responsive transit and traditional buses. *Int. J. Transp. Sci. Technol.* **2024**, *15*, 81–101. [CrossRef]
18. Abduljabbar, R.; Dia, H.; Liyanage, S.; Bagloee, S.A. Applications of Artificial Intelligence in Transport: An Overview. *Sustainability* **2019**, *11*, 189. [CrossRef]
19. 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**, *2677*, 605–623. [CrossRef]
20. Mohammadnazar, A.; Arvin, R.; Khattak, A.J. Classifying travelers' driving style using basic safety messages generated by connected vehicles: Application of unsupervised machine learning. *Transp. Res. Part. C Emerg. Technol.* **2021**, *122*, 102917. [CrossRef]
21. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]
22. Swana, E.; Doorsamy, W. An Unsupervised Learning Approach to Condition Assessment on a Wound-Rotor Induction Generator. *Energies* **2021**, *14*, 602. [CrossRef]
23. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
24. 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]
25. Darwish, A.M.; Almansour, M.; Salah, A.; Zagow, M.; Saeed, K.; Elkafoury, A. Sensitivity evaluation of machine learning-based calibrated transportation mode choice models: A case study of Alexandria City, Egypt. *Transp. Res. Interdiscip. Perspect.* **2024**, *24*, 101052. [CrossRef]
26. Pineda-Jaramillo, J.; Arbeláez-Arenas, Ó. 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]
27. Liu, J.; Wang, B.; Xiao, L. Non-linear associations between built environment and active travel for working and shopping: An extreme gradient boosting approach. *J. Transp. Geogr.* **2021**, *92*, 103034. [CrossRef]
28. García-Cerrud, C.A.; Flores De La Mota, I. Methodology for public passenger transport in developing countries: A survey. *Transp. Res. Procedia* **2023**, *71*, 148–155. [CrossRef]
29. Kazemeini, A.; Kermanshah, A. Promoting Sustainable Transport in Developing Countries: A Case Study of University Students in Tehran. *Future Transp.* **2023**, *3*, 858–877. [CrossRef]
30. Sánchez-Atondo, A.; García, L.; Calderón-Ramírez, J.; Gutiérrez-Moreno, J.M.; Mungaray-Moctezuma, A. Understanding Public Transport Ridership in Developing Countries to Promote Sustainable Urban Mobility: A Case Study of Mexicali, Mexico. *Sustainability* **2020**, *12*, 3266. [CrossRef]
31. Cervero, R. Linking urban transport and land use in developing countries. *J. Transp. Land Use* **2013**, *6*, 7–24. [CrossRef]
32. Elkafoury, A.; Elboshy, B.; Darwish, A.M. Development of response surface method prediction model for traffic-related roadside noise levels based on traffic characteristics. *Environ. Sci. Pollut. Res.* **2023**, *30*, 94229–94241. [CrossRef] [PubMed]
33. Vythoulkas, P.C.; Koutsopoulos, H.N. Modeling discrete choice behavior using concepts from fuzzy set theory, approximate reasoning and neural networks. *Transp. Res. Part. C Emerg. Technol.* **2003**, *11*, 51–73. [CrossRef]
34. Lee, D.; Mulrow, J.; Haboucha, C.J.; Derrible, S.; Shiftan, Y. Attitudes on Autonomous Vehicle Adoption using Interpretable Gradient Boosting Machine. *Transp. Res. Rec.* **2019**, *2673*, 865–878. [CrossRef]
35. Biagioni, J.P.; Szczurek, P.M.; Nelson, P.C.; Abolfazl, M. Tour-Based Mode Choice Modeling: Using an Ensemble of Conditional and Unconditional Data Mining Classifiers. In Proceedings of the Transportation Research Board 88th Annual Meeting, Washington, DC, USA, 11–15 January 2009.
36. Omrani, H. Predicting Travel Mode of Individuals by Machine Learning. *Transp. Res. Procedia* **2015**, *10*, 840–849. [CrossRef]
37. Richards, M.J.; Zill, J.C. Modelling mode choice with machine learning algorithms. In Proceedings of the Australasian Transport Research Forum 2019, Canberra, Australia, 30 September–2 October 2019.
38. Hagenauer, J.; Helbich, M. A comparative study of machine learning classifiers for modeling travel mode choice. *Expert. Syst. Appl.* **2017**, *78*, 273–282. [CrossRef]
39. Abulibdeh, A. Analysis of mode choice affects from the introduction of Doha Metro using machine learning and statistical analysis. *Transp. Res. Interdiscip. Perspect.* **2023**, *20*, 100852. [CrossRef]
40. Omrani, H.; Gerber, P.; Charif, O.; Awasthi, A.; Trigano, P. Prediction of individual travel mode with evidential neural network model. *Transp. Res. Rec.* **2013**, *2399*, 1–8. [CrossRef]
41. Xie, C.; Lu, J.; Parkany, E. Work Travel Mode Choice Modeling with Data Mining: Decision Trees and Neural Networks. *Transp. Res. Rec.* **2003**, *1854*, 50–61. [CrossRef]
42. Hensher, D.A.; Ton, T.T. A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. *Transp. Res. E Logist. Transp. Rev.* **2000**, *36*, 155–172. [CrossRef]

43. Zhang, Y.; Xie, Y. Travel Mode Choice Modeling with Support Vector Machines. *Transp. Res. Rec. J. Transp. Res. Board.* **2008**, *2076*, 141–150. [CrossRef]
44. Sekhar, C.R.; Minal Madhu, E. Mode Choice Analysis Using Random Forrest Decision Trees. *Transp. Res. Procedia* **2016**, *17*, 644–652. [CrossRef]
45. 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]
46. Ding, C.; Cao, X.; Næss, P. Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. *Transp. Res. Part. A Policy Pract.* **2018**, *110*, 107–117. [CrossRef]
47. Ding, C.; Wu, X.; Yu, G.; Wang, Y. A gradient boosting logit model to investigate driver's stop-or-run behavior at signalized intersections using high-resolution traffic data. *Transp. Res. Part. C Emerg. Technol.* **2016**, *72*, 225–238. [CrossRef]
48. Tan, L.; Ma, C.; Xu, X.; Xu, J. Choice Behavior of Autonomous Vehicles Based on Logistic Models. *Sustainability* **2019**, *12*, 54. [CrossRef]
49. Haboucha, C.J.; Ishaq, R.; Shiftan, Y. User preferences regarding autonomous vehicles. *Transp. Res. Part. C Emerg. Technol.* **2017**, *78*, 37–49. [CrossRef]
50. Elbany, M. Autonomous Vehicles Probability-to-Choose and Public Usage Behavior Study in Developing Countries; Case Study: Port Said City. *Port-Said Eng. Res. J.* **2024**, *28*, 50–63. [CrossRef]
51. Darwish. Development of Transport Mode Choice Model between Public and Private Transport in Densely Metropolitan Areas, Alexandria as a Case Study. Ph.D. Thesis, Alexandria University, Alexandria, Egypt, 2021.
52. Elharoun, M.; Shahdah, U.E.; El-Badawy, S.M. Developing a Mode Choice Model for Mansoura city in Egypt. *Int. J. Traffic Transp. Eng.* **2018**, *8*, 528–542. [CrossRef]
53. El Esawey, M.; Ghareib, A. Analysis of Mode Choice Behavior in Greater Cairo Region. In Proceedings of the Transportation Research Board 88th Annual Meeting, Washington, DC, USA, 11–15 January 2009.
54. Abdel-Aal, M.M.M. Value of time determination for the city of Alexandria based on a disaggregate binary mode choice model. *Alex. Eng. J.* **2017**, *56*, 567–578. [CrossRef]
55. El-Bany, M.E.S.; Shahin, M.M.; Hashim, I.H.; Serag, M.S. Policy sensitive mode choice analysis of Port-Said City, Egypt. *Alex. Eng. J.* **2014**, *53*, 891–901. [CrossRef]
56. Li, X.; Shi, L.; Shi, Y.; Tang, J.; Zhao, P.; Wang, Y.; Chen, J. Exploring interactive and nonlinear effects of key factors on intercity travel mode choice using XGBoost. *Appl. Geogr.* **2024**, *166*, 103264. [CrossRef]
57. Wang, Y.; Pan, Z.; Zheng, J.; Qian, L.; Li, M. A hybrid ensemble method for pulsar candidate classification. *Astrophys. Space Sci.* **2019**, *364*, 1–13. [CrossRef]
58. Assele, S.Y.; Meulders, M.; Vandebroek, M. Sample size selection for discrete choice experiments using design features. *J. Choice Model.* **2023**, *49*, 100436. [CrossRef]
59. Yin, C.; Wu, J.; Sun, X.; Meng, Z.; Lee, C. Road transportation emission prediction and policy formulation: Machine learning model analysis. *Transp. Res. D Transp. Environ.* **2024**, *135*, 104390. [CrossRef]
60. Cai, J.; Wang, M.; Wu, Y. Research on Pedestrian Crossing Decision Models and Predictions Based on Machine Learning. *Sensors* **2024**, *24*, 258. [CrossRef] [PubMed]
61. Hillel, T.; Bierlaire, M.; Elshafie, M.Z.E.B.; Jin, Y. A systematic review of machine learning classification methodologies for modelling passenger mode choice. *J. Choice Model.* **2021**, *38*, 100221. [CrossRef]
62. Bhuiya, M.M.R.; Hasan, M.M.U.; Keellings, D.J.; Mohiuddin, H. Application of Machine Learning Classifiers for Mode Choice Modeling for Movement-Challenged Persons. *Future Transp.* **2022**, *2*, 328–346. [CrossRef]
63. Kulkarni, A.; Chong, D.; Batarseh, F.A. Foundations of data imbalance and solutions for a data democracy. *Data Democr. Nexus Artif. Intell. Softw. Dev. Knowl. Eng.* **2020**, 83–106. [CrossRef]
64. Montesinos López, O.A.; Montesinos López, A.; Crossa, J. Overfitting, Model Tuning, and Evaluation of Prediction Performance. In *Multivariate Statistical Machine Learning Methods for Genomic Prediction*; Springer: Cham, Switzerland, 2022; pp. 109–139. [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.