

Article

Understanding and Predicting the Usage of Shared Electric Scooter Services on University Campuses

Seyed Mohammad Hossein Moosavi ^{1,*}, Zhenliang Ma ², Danial Jahed Armaghani ^{3,*}, Mahdi Aghaabbasi ⁴, Mogana Darshini Ganggayah ⁵, Yuen Choon Wah ¹ and Dmitrii Vladimirovich Ulrikh ³

¹ Department of Civil Engineering, Faculty of Engineering, Universiti of Malaya, Kuala Lumpur 50603, Malaysia

² Department of Civil and Architectural Engineering, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden

³ Department of Urban Planning, Engineering Networks and Systems, Institute of Architecture and Construction, South Ural State University, 76 Lenin Prospect, 454080 Chelyabinsk, Russia

⁴ Transportation Institute, Chulalongkorn University, Bangkok 10330, Thailand

⁵ Malaysia School of Business, Monash University, Subang Jaya 47500, Selangor, Malaysia

* Correspondence: mh.moosavi65@gmail.com (S.M.H.M.); danialarmaghani@susu.ru (D.J.A.)

Abstract: Electric vehicles (EVs) have been progressing rapidly in urban transport systems given their potential in reducing emissions and energy consumptions. The Shared Free-Floating Electric Scooter (SFFES) is an emerging EV publicized to address the first-/last-mile problem in travel. It also offers alternatives for short-distance journeys using cars or ride-hailing services. However, very few SFFES studies have been carried out in developing countries and for university populations. Currently, many universities are facing an increased number of short-distance private car travels on campus. The study is designed to explore the attitudes and perceptions of students and staff towards SFFES usage on campus and the corresponding influencing factors. Three machine learning models were used to predict SFFES usage. Eleven important factors for using SFFESs on campus were identified via the supervised and unsupervised feature selection techniques, with the top three factors being daily travel mode, road features (e.g., green spaces) and age. The random forest model showed the highest accuracy in predicting the usage frequency of SFFESs (93.5%) using the selected 11 variables. A simulation-based optimization analysis was further conducted to discover the characterization of SFFES users, barriers/benefits of using SFFESs and safety concerns.

Keywords: green campus; shared free-floating electric scooter; usage frequency prediction; decision tree; random forest



Citation: Moosavi, S.M.H.; Ma, Z.; Armaghani, D.J.; Aghaabbasi, M.; Ganggayah, M.D.; Wah, Y.C.; Ulrikh, D.V. Understanding and Predicting the Usage of Shared Electric Scooter Services on University Campuses. *Appl. Sci.* **2022**, *12*, 9392. <https://doi.org/10.3390/app12189392>

Academic Editor: Mirco Peron

Received: 2 August 2022

Accepted: 13 September 2022

Published: 19 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The advancement of information technology and sharing economy business models is changing traditional models of ownership and transport services. New modes of travel are emerging in urban areas, such as transport network company services, bike-sharing and scooter-sharing, etc. Shared micro-mobility (SMM, the shared utilization of an e-/bicycle, e-/scooter, or other low-speed modes) is a newly developed transportation mode [1]. SMM provides users with a short-term access to a transportation service over an as-needed basis [2].

Early documented impacts of SMM include increased mobility [3], decreased greenhouse gas emissions [4], and decreased automobile use [5,6]. Since 2017, over USD 5.7 billion have been devoted to SMM start-up companies, mostly in China. A steady customer pool has been established in the SMM market, which is two to three times faster than ride-hailing or car-sharing services. The combined value of SMM start-ups is estimated to exceed USD 1 billion [7].

Shared Free-Floating Electric Scooters (SFFESs) have been altering travel in cities and on university campuses. Though SFFESs have swiftly obtained popularity and approval over the past couple of years, limited studies have been reported on their use. The analysis of Berg Insight shows that the COVID-19 crisis led to a lower shared-scooter ridership in 2021. However, in the long term, ridership is projected to reach over 4.6 million people in 2024 worldwide, with a base of 774,000 people in 2019 [8].

New mobility services, such as Uber/Lyft, have been changing the landscape of urban mobility. SFFESs have become increasingly popular and utilized by communities given their acceptable cost, zero-emission power and minimal environmental footprint. In addition, given the present pandemic, city planners are looking for new methods, such as SFFESs, to reconcile urban mobility need and social distance requirements. While SFFESs offer promising opportunities, they also bring negative externalities, including safety and equity issues for pedestrians, cyclists and disabled/elder citizens [6,9]. Many cities proscribed SFFES services, particularly in the initial boom of SMM, due to vital vandalism and street clutter, including Austin, Nantes, Amsterdam, Bordeaux, and recently Kuala Lumpur. These cities revisited their decisions afterwards and devised novel regulation provisions to optimize SFFES benefits while limiting their drawbacks. Some cities banned the usage of SFFESs due to regulation requirements; for example, the New York State Department of Motor Vehicles legislation requires the registration of any electric vehicle, which makes SFFES service impossible.

Effective regulation faces two major setbacks emanating from the organizational culture/climate mismatch between local authorities and service operators. The operators need a high vehicle density to guarantee a high service quality and ultimately foster their market [10,11]. However, local authorities are wary of street clutter and intend to limit the fleet size. While technology and investments are essential for service implementation, equally important is to signify the impact of shared micro-mobility on the urban mobility ecosystem and its evolution trend over time, in order to better design and integrate it into sustainable mobility as a whole [12]. However, very few studies examined SFFESs, and in addition the existing studies were limited in the analysis approaches used, which may fail to capture the complex nonlinear relationship between variables. In addition, most studies on SFFES services were conducted in the United States, China and, most recently, European cities, but are yet very limited in developing countries. It is presumed that this study will be the first step taken to assess the adoption of SFFESs and usage behavior with respect to a Malaysian context. The paper identifies public concerns, SFFES benefits and barriers, and the choice and usage behavior of the university population (students and non-/academic staff).

Choice behavior in new mobility services is usually assessed and modeled using traditional statistical models, such as regression, mixed logit, multinomial and binary logit models [13,14]. Recently, [15] used the Chi² and Kruskal–Wallis tests to analyze the frequency of e-scooter use. Given the strict assumptions of statistical models, they have limited capabilities to capture the complex relationships between factors and choices, nonlinear correlations among factors, and to deal with factors with various categories [16]. Machine learning (ML) methods have been widely utilized in civil engineering [17–23] and transportation studies [24,25]. They can model the nonlinear associations between independent and target variables as well as among independent variables [26,27]. Therefore, it can be argued that the current study is one of the first attempts to predict SFFES usage frequency and identify significant factors impacting its use by adopting ML techniques.

Malaysian universities are currently adopting new sustainable strategies in moving their campuses towards becoming green campuses. Specifically, the management of the University of Malaya is planning to launch SFFES service in the near future. This paper aims to predict the usage frequency of SFFESs among the students and staff on the campus. As a summary of above discussion, the main contributions are:

1. This research study is one of the first efforts made to scrutinize the usage of SFFESs on a large university campus. In addition, this is perhaps the first study on SFFES services in “developing countries” such as Malaysia.

2. This research is one of the first studies which aims to predict the usage frequency of SFFESs and pinpoint significant attributes affecting the use of SFFESs by adopting various supervised and unsupervised machine learning techniques.

The remaining of the article is organized as follows: Section 2 concerns a literature review on related works, followed by the survey design and data collection in Section 3. Section 4 proposes the analysis methodology, including feature selection and model development. Sections 5 and 6 present the model output, analysis results and simulation-based optimization and discussion. The final part offers the obtained findings of the study and suggests future directions.

2. Related Works

It is believed that, in terms of urban features and population, higher education organizations mirror smaller cities [28]. Moreover, there exist many activities occurring on university campuses that exert both direct and indirect effects on the natural milieu [29]. Therefore, practitioners in these academic contexts need to apply green practices and provide support in offering multidisciplinary green technical solutions to achieve sustainable development on campuses [30]. The United States Green Building Council [31] revealed that a green campus is a higher education community seeking to enhance its resource conservation, energy efficiency, and ecological quality via training on healthy living, sustainability, and convenience learning environments for all.

In the context of higher education, green practices are rising rapidly. However, achieving sustainability in Malaysian universities in this regard has yet remained an issue [32]. Malaysia has committed itself to buttress sustainability on university campuses after signing the Talloires Declaration. Thereafter, enthusiasm for focusing more on sustainable development has increased in Malaysia. Nevertheless, many universities yet lag behind in green practices in order to attain sustainability as an institutional policy. This runs counter to the outline of higher education institutions since the 1992 Earth Summit in Rio. Universities are facing pressure from non-/governmental organizations to incorporate green practices in their activities following several sustainability declarations.

Shared micro-mobility (SMM)—the short-term rent of micro-mobility vehicles such as (e-)scooters and (e-)bicycles—is regarded as a mobility (sub)system that can alter the present transport system in terms of cars [2,33]. This technology was first presented in 2017 and has now become an important mode of transport emerging in more than 1000 cities and college campuses worldwide. Such web-based SFFES services are managed by rental networks and operated using smartphones.

Academic studies on SFFESs have been emerging. For example, ref. [34] examined anonymized SFFES trip data and concluded that users ride SFFESs for about 8 min for 0.7 miles, with an average speed of 5.23 miles per hour. The SFFES service could be used as an appropriate travel model for last-mile transport or short-distance trips. Ref. [35] found considerable differences in temporal and spatial usage patterns between SFFESs and docked bike-sharing ridership [36]. Ref. [37] assessed the behavioral determinants of travelers’ purpose for using SFFESs and found that the perceived compatibility of SFFESs significantly affected usage intention. Ref. [15] performed Kruskal–Wallis and Chi² test with e-scooter-associated survey data and pointed to the importance of sociodemographic characteristics in affecting SFFES usage. Ref. [38] evaluated the API data of SFFES vendors and found significant SFFES ridership variations between weekends and weekdays, but not between morning and afternoon trips.

As mentioned before, most of the academic studies in this field were conducted in the US. Surveys were conducted by a few cities to complement assessing the e-scooter pilot programs. It was found that e-scooters were popular or generally considered to present a respected service, even among non-users [15]. For example, the Portland report

stated that over 30% of people had tried e-scooters. Over 70% of Portlanders riding an e-scooter stated that they utilized e-scooters most commonly for transportation, but not recreation. The reasons for use included reliability, speed, cost, convenience and fun [39]. Unequal adoptions between population groups were suggested by surveys. The gender (female/male) splits were 64/34 and 70/30 for Portland and Denver. In total, 69% of e-scooter users were aged 20–39 in Portland, while the figure was over 50% in Denver [39,40].

An online survey was performed by the [41] on 1250 individuals in the five largest cities of Germany (Hamburg, Berlin, Frankfurt, Cologne, and Munich) in September 2019. It explored their overall mobility behavior and utilization of SFFES systems. It revealed that 42.7% of e-scooter users were aged between 18–25, and 28.8% between 26–35. The SFFES service substituted 49.1% of walking trips and 64.5% of public transport trips. A quantitative study was performed by the French [42]. They gathered 4382 user responses after various semi-structured and exploratory interviews. It reported that e-scooter renters were young (52% younger than 34), male (66%), highly educated (19% students, 53% work executives), and with a significant share of non-locals (42%). For the modal shift, users substituted walking (44%), public transport (30%), and bike trips (3% owned a bike; 9% shared a bike).

Supervised learning algorithms learn correlation patterns from data (independent and target variables) and make decisions/predictions based on a specific objective. Decision trees (DT) are widely used in data-driven prediction analysis [43–46]. Decision trees have been used for model evaluation and identifying important variables. Random forests (RF), a derivation of decision trees, can work in both supervised and unsupervised modes. It can handle continuous as well as categorical data in classification or regression tasks [47,48]. Random forests are prioritized over other techniques, as it can manage highly non-linear data, and demonstrates many features, such as agility in locating noise in data and adjustability to parameters [49]. It has three main features: (i) estimating missing values automatically, (ii) Weighted Random Forest (WRF) for balancing errors in imbalanced data, and (iii) estimation of the significance of variables utilized for categorization [50]. Naïve Bayes (NB) classifiers are also able to handle continuous and categorical variables and quickly make real-time predictions [51].

Unsupervised learning is designed to analyze unlabeled data [52]. As the amount of unlabeled data is exponentially rising, it is essential to explore unsupervised learning to perform feature selection. Data clustering (feature selection) is an important problem in knowledge discovery to improve the understandability, scalability and accuracy of resulting models. The clusters correspond to hidden models and the resulting outcomes represent data notions. In the context of supervised learning, feature selection refers to predictions based on provided outputs, while in unsupervised learning the features are clustered without any prior knowledge of the expected output. The importance of feature clustering is to improve prediction performance and provide a deeper understanding of the underlying process that produces the data. Examples of clustering algorithms are k-means, partitioning around medoids (PAM) and hierarchical clustering. This paper uses both supervised and unsupervised learning techniques for feature selections and predicting the usage frequency of SFFESs on campus.

3. Methodology

3.1. Survey Design and Data Collection

The survey was designed to understand the adoption, choice behaviour and usage of SFFES services on the university campus. The questionnaire consisted of 55 mandatory questions covering the following aspects:

- Sociodemographic information, including information about age, gender, marital status, residential area, highest level of education, employment status, race, household monthly income, private vehicle ownership, shared mobility and membership and frequency of usage of e-hailing services.

- Commuting characteristics, including commuting mode to and from the campus, and the travel mode, frequency, distance, time and cost on campus.
- Perceptions and choices regarding the SFFES service, including (1) perceptions regarding using SFFESs and concerns of safety, equity, costs, comfort, and social distancing due to COVID-19; (2) service attributes, such as accessibility, payment methods, and the advantages and disadvantages of shared e-scooters compared to other transport modes; and (3) infrastructure and built environment, such as separated lanes for scooters, green spaces, quality of road surfaces and connectivity.
- Usage frequency of the SFFES service, including four levels of response: (1) not using an e-scooter at all; (2) using an e-scooter as a mode of transport occasionally (sometimes but infrequently); (3) using an e-scooter frequently; and (4) using an e-scooter regularly as a main mode of transport. (Table 1 presents the information on the data and attributes used in this study).

Table 1. Variables used in this study for analysis.

Attribute	Description	Values
Sociodemographic		
Age	Age	(1) 18 to 29; (2) 30 to 44; (3) 45 to 60; (4) Over 60
Gender	Gender	(1) Male; (2) Female
Education	Highest education level	(1) Secondary; (2) Diploma; (3) Bachelor's degree; (4) Master's degree; (5) Doctorate degree
Position	Job position	(1) Undergraduate student; (2) Postgraduate student; (3) Academic staff; (4) Non-academic staff
Status	Employment/education status	(1) Full-time; (2) Part-time
Race	Race	(1) Chinese; (2) Malay; (3) Indian; (4) Other
Monthly Income	Monthly household income	(1) Less than RM 2000; (2) Between RM 2000 RM 4000; (3) Between RM 4000 and RM 6000; (4) Between RM 6000 and RM 12,000; (5) More than RM 12,000
Private vehicle	Private vehicle ownership	(1) Yes; (2) No
E-hailing	Usage of e-hailing services per week	(1) Not using at all; (2) Less than 3 times; (3) 3 to 6 times; (4) More than 6 times
SMS Membership	Membership of shared mobility services	(1) Yes; (2) No
Travel characterization		
Travel mode	Usual travel mode for going to campus	(1) E-hailing taxi; (2) Private car; (3) Private motorcycle; (4) Public transportation; (5) Walking/cycling
Camp.Hrs/d	Hours usually spent on the campus per day	(1) 1 to 3 h; (2) 3 to 5 h; (3) 5 to 8 h; (4) More than 8 h
Camp.Tra/d	Number of journeys onto or to outside of the campus per day	(1) Less than 2 journeys; (2) 2 to 4 journeys; (3) 4 to 6 journeys; (4) More than 6 journeys
Camp.mod/d	Travel mode on the campus	(1) E-hailing taxi; (2) Private car; (3) Private motorcycle; (4) Public transportation; (5) Walking/cycling
Camp.tra.time/d	Duration of daily travel on the campus	(1) Less than 10 min; (2) 10 to 20 min; (3) 20 to 30 min; (4) More than 30 min
Camp.tra.cost/d	Daily travel cost on the campus	(1) Less than RM 5; (2) Between RM 5 and RM15; (3) Between RM15 and RM 25; (4) More than RM25
Attitudinal factors: impact of infrastructure		
Sep.lane	Bike/scooter lane separate from road traffic	(1) Strongly discourage; (2) Discourage; (3) Encourage; (4) Strongly encourage
On-road.Lane	Bike/scooter lane on the road with traffic	(1) Strongly discourage; (2) Discourage; (3) Encourage; (4) Strongly encourage
No-Lane	Road with no bike/scooter lane	(1) Strongly discourage; (2) Discourage; (3) Encourage; (4) Strongly encourage
Greenery	Green Space (e.g., road-side trees, greenery, water)	(1) Strongly discourage; (2) Discourage; (3) Encourage; (4) Strongly encourage
Smooth.Surf	A smooth road surface	(1) Strongly discourage; (2) Discourage; (3) Encourage; (4) Strongly encourage
Connectivity	Pathways/roads connectivity	(1) Strongly discourage; (2) Discourage; (3) Encourage; (4) Strongly encourage
e-scooter Usage (Target variable)	Shared e-scooter frequency of usage	(1) Not using at all; (2) Sometimes/infrequently; (3) Frequently; (4) Regularly as the main mode of transport.

The survey was carried out on students and staff of the University of Malaya (UM). The UM is situated in the southwest of Kuala Lumpur. It has a 373.12-hectare campus and houses around 20,000 students and 6000 staff. In addition to these numbers, many

daily operations, activities, and events require continuous mobility access to different transportation modes. Consequently, integrated transportation system management on the university campus is pivotal. The current transportation services on the UM campus include bus services (campus and traditional buses), a bicycling facility, and car and pedestrian accessibility. Figure 1 shows the University Campus Map and the road line map.

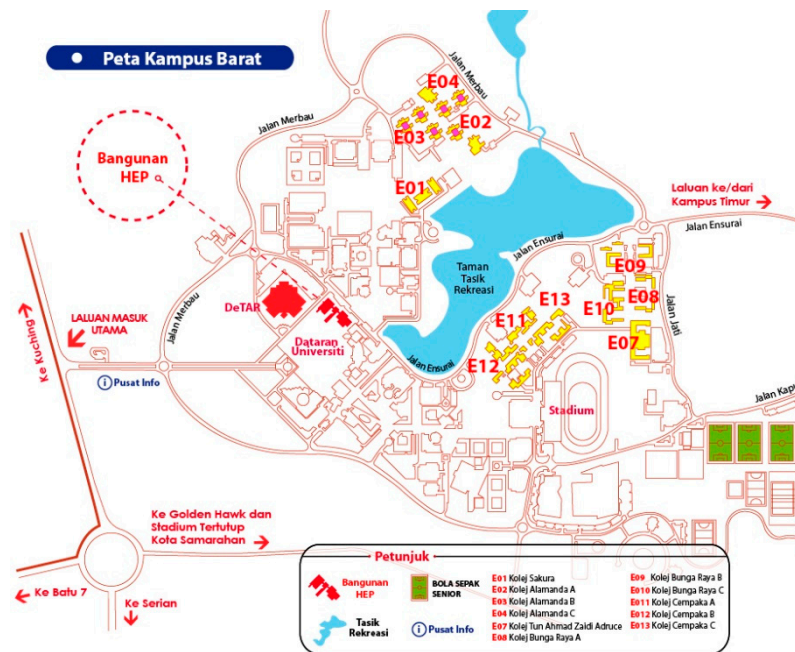


Figure 1. University of Malaya campus map.

The online Google questionnaire survey was disseminated to over 30,000 faculty, non-academic staff and university students in December 2020. The survey was estimated to take 10 min to complete. The survey link was active for a period of three weeks. We received 1023 responses and 1000 surveys were valid for further analysis (response rate: 1.7%).

Table 2 captures the sociodemographic characteristics of the sample, the UM population and the overall university populations in Malaysia. For the university population, we used the data statistics of 2020. The gender distribution in the sample is overall representative, with the female population slightly overrepresented. Shares for occupation composition are comparable. Given the similarities of gender and occupation, we believe that the sample sufficiently reflects the socioeconomic features of the targeted population.

Table 2. Comparison between the survey sample and the university population in percentage.

Socio-Demographics	Total Sample (n = 1000)	UM University	All Universities in Malaysia
<i>Gender</i>			
Male	45.6	49.0	47.0
Female	54.4	51.0	53.0
<i>Occupation</i>			
Undergraduate students	51.5	51.7	48.5
Graduate students	36.5	27.6	33.5
Part-time graduate students	2.1	6.3	6.3
Faculty and staff	9.9	16.3	11.7

3.2. Feature Selection

Feature selection is an option in statistics to detect significant factors that use measures of confidence intervals as well as hypothesis testing. After conducting model evaluation, the elements (independent variables) must be examined further to see how they lead to measurement accuracy. Hence, machine learning algorithms are built-in with the feature selection technique to analyze the variables or features in the input data. The distribution of these features contributes to the prediction of the final outcome using machine learning models. Feature selection helps to understand the model better by focusing only on the important variables. This statistical technique eliminates variables which are insignificant or highly correlated with any other variable. Based on significance score, the order of variables can be illustrated to realize the accuracy of prediction. The reliability of important variables depends on the accuracy of a specific algorithm. The objectives of feature selection in machine learning are to reduce the complexity of the mode and to promote the performance of the model. Feature selection evaluates the relationship between the input variables and target variable.

The supervised and unsupervised feature selection methods vary considering the target variables. While the supervised learning model requires a target variable to specify the important variables, the unsupervised learning model disregards the target variable and chooses important variables using correlation. Figure 2 shows the study methodology workflow.

Clustering

The unsupervised learning model clusters the input variables based on correlation between each other, and without considering the target variable. The important variables obtained from the random forest feature selection are used to perform clustering. There are two steps in clustering: (a) determination of the optimal number of clusters, and (b) hierarchical clustering.

To determine the optimal number of clusters:

The optimal number of clusters is specified using the Gap Statistics method. The *fviz_nbclust()* function in *factoextra* R package is employed to compute the optimal number of clusters. The Gap Statistics algorithm works as follows [53]:

- The observed data of 1000 samples with n variables is analyzed by changing the number of clusters from $k = 1, \dots, k_{max}$, and the total within intra-cluster variation W_k is computed.
- B reference datasets with a random uniform distribution is generated. Each reference dataset is clustered with varied number of clusters $k = 1, \dots, k_{max}$, and the corresponding total within intra-cluster variation W_{kb} is computed.
- The estimated gap statistic is computed as the deviation of the observed W_k value from its expected value, W_{kb} under the null hypothesis: $Gap(k) = 1B \sum b = 1B \log(W * kb) - \log(Wk)$. The standard deviation of the statistics is also computed.
- The number of clusters is chosen as the smallest value of k such that the gap statistic is within one standard deviation of the gap at $k + 1$: $Gap(k) \geq Gap(k + 1) - sk + 1$.

The optimal number of clusters is used to perform hierarchical clustering using the *hclust* R package. Hierarchical clustering is an agglomerative clustering algorithm, which can be cut at a chosen height to produce the desired number of clusters [54,55]. The clusters produced in a dendrogram are joined together in order of their closeness measured by dissimilarity. The steps of hierarchical clustering are as follows:

- Divide n variables into k groups by cutting at a desired similarity level.
- Calculate the dissimilarity matrix between variables using function *dist ()* in *hclust* package.
- Plot the dendrogram using *fviz_dend ()* function in *factoextra* package with dissimilarity matrix as the input.

Correlation analysis is performed using R *corrplot* function to assess the relationship between the variables.

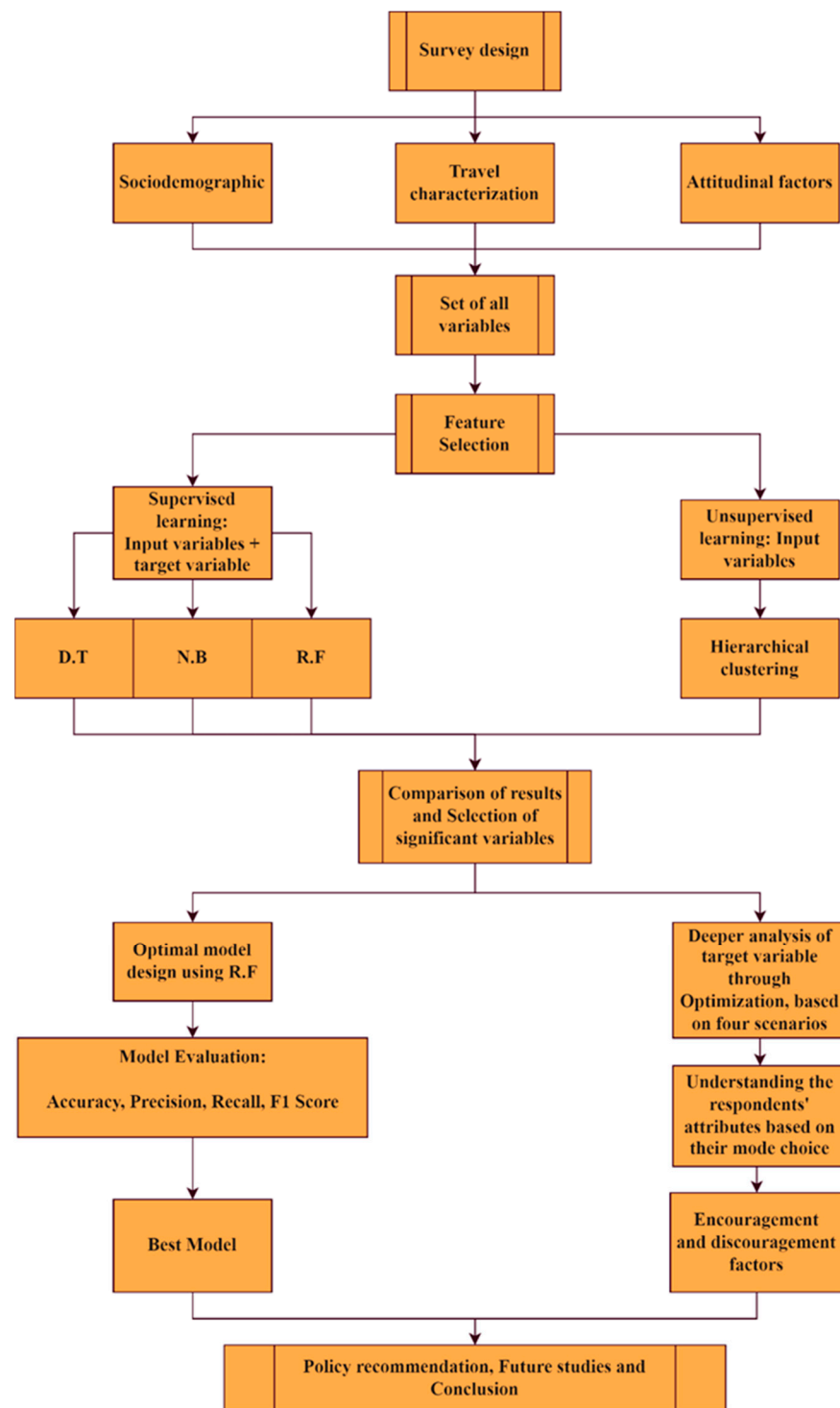


Figure 2. Methodology workflow.

3.3. The Optimal Model Design

The model assessment is performed using the important variables selected through supervised (random forest, decision tree and Naïve Bayes) and unsupervised learning methods. After selection of significant variables, the random forest classifier is used to assess the model performance using the test and out-of-bag errors by changing the total number of trees (*ntree*) and predictors at each split (*mtry*). The best *ntree* and *mtry* are obtained using the measures of the mean squared error and variance, calculated using the out-of-bag errors. A total of 2/3 of the data is used for training and 1/3 for validating the

trees. The final model is developed using the best *ntree* and *mtry*. Random forest algorithm is a joint and collaborative learning algorithm that is derived from decision trees. It follows the rules of decision trees but constructs numerous decision trees during training time and outputs the class with maximum vote. For example, the random forest algorithm constructs trees of different classes using the similar input data. The tree structures can be explained using subset matrices as shown in Figure 3. Three random subsets are created during the training process. Three different trees are explained using three subsets (S1, S2 and S3). Different samples are grouped into different subsets based on the correlation between input features (independent variables). Decision trees are built based on the subset values. The decisions or the final predicted output from each decision tree is considered a class. The class, which receives maximum votes from the total number of trees, will be chosen as the final output. Class 1 has two votes whereas class 2 has one vote in Figure 3, therefore class 1 is the final predicted output. This class 1 will be used to rank the variables based on importance score.

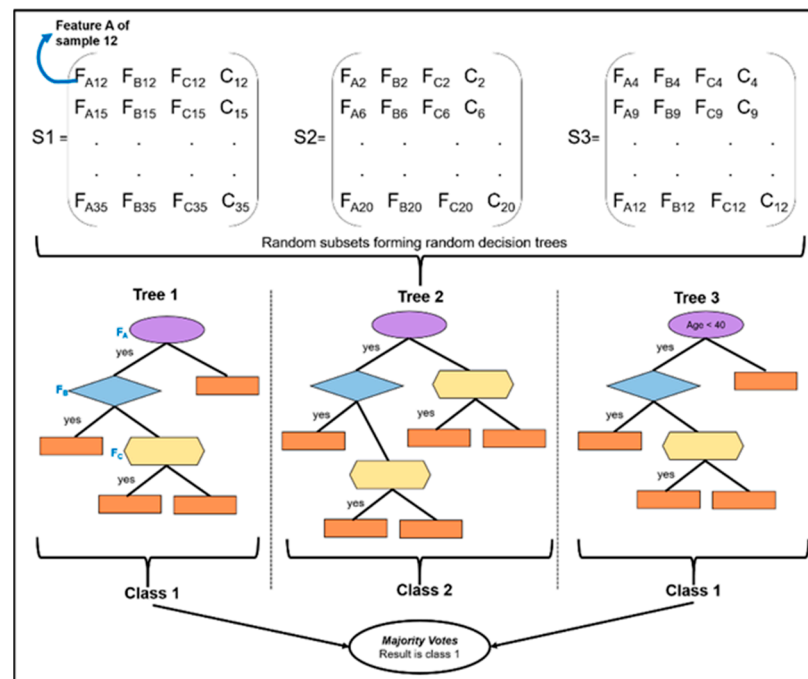


Figure 3. Random forest algorithm workflow.

The R package *random-forest* is used to perform feature selection for n variables, and the number of important variables is determined in three stages: initial, threshold and prediction. The most important variables are selected from the final prediction stage. Random forest considers a random subset of predictors, p , each time when splitting the training set. The trees find all the predictors while performing a split and select the best amongst them. The total number of predictors at each split is calculated using the formula $mtry = \sqrt{n}$. The default number of trees used in random forest feature selection is $ntree = 500$ and the total number of predictors used to construct the trees is \sqrt{n} .

3.4. Model Evaluation

Model evaluation in machine learning is an alternative to the assessment of effect size in conventional statistics [56]. It is a key step in machine learning, as the ability of the model to make predictions on unseen or future samples will enhance the trust on the model to be used in a particular dataset. The measurement for model evaluation is accuracy in percentage (estimate of generalization of a model on future data). The most popular model evaluation technique is cross-validation. Cross-validation divides the data into test (independent dataset) and training (subset of data used to train the model for future

predictions) sets; 5-fold cross-validation was performed. The accuracy is assessed based on the overall error estimation comparing the test and training sets. An interchange of test and training sets reduces bias and variance in the method. Cross-validation can be used to compare the performance of different machine-learning algorithms on the same data, as this will make it easier to select the best algorithm to perform further analyses. A confusion matrix is the most common interpretation of model performance in supervised learning. A confusion matrix can produce model accuracy, precision, recall and F1 score. In this study, the total number of samples ($n = 1000$) was divided into 80% of the training set and 20% of the testing set. The model evaluation was performed using three different algorithms: decision tree, random forest and Naïve Bayes, and the accuracy measures based on the confusion matrix were recorded.

4. Results

4.1. Descriptive Analysis (Encouragement and Discouragement Factors)

This section presents the results of the last part of the survey, which measured encouragement and discouragement factors for using SFFESs. In other words, after predictions of SFFES usage, important factors and levels of acceptance between different groups of respondents, this section was designed to answer the following questions: 1—Why will certain respondents never use SFFESs (8% of total respondents according to Figure 4)? 2—What are their main concerns? 3—What are the benefits of the SFFES service from our respondents' point of view?

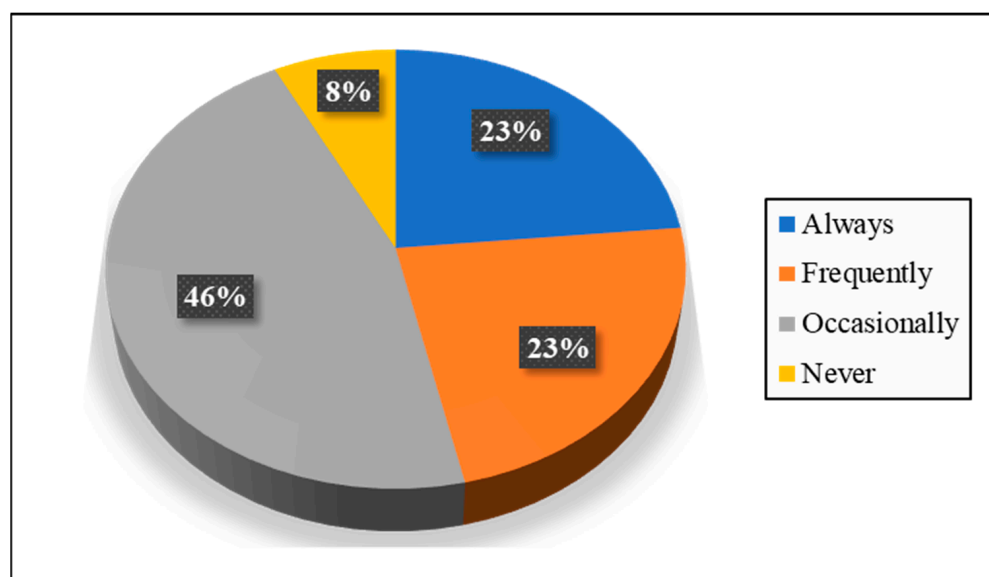


Figure 4. Percentage of SFFES use frequency based on four categories.

Survey participants were asked to express their perceptions and feelings regarding the encouragement and discouragement factors of using SFFES services. In the first part, we asked the participants about the benefits and advantages of SFFESs. Figure 5 presents the overall responses to questions about the benefits/advantages of using SFFESs.

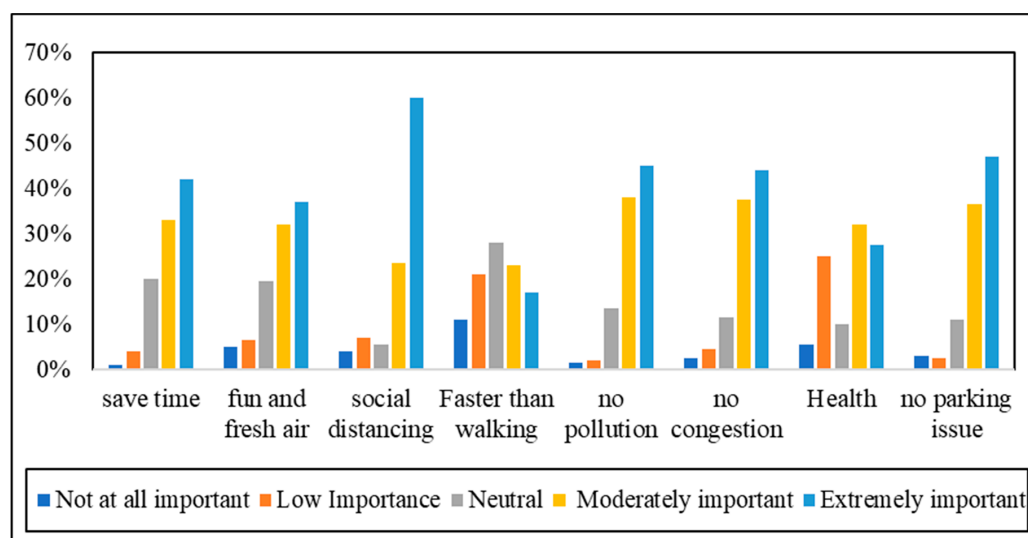


Figure 5. Participants’ perceptions about the advantages/benefits of using SFFESs.

According to Figure 5, most of the respondents were almost agreed on all mentioned benefits of SFFESs, except for “speed” and “physical/mental health”. A considerable number of participants used a private vehicle as their main mode of transport on the campus. This could be one of the reasons why most of the participants selected a neutral option about the speed. The second, less-important advantage of the SFFES was health benefits, based on participants’ responses. Indeed, the physical and mental benefits of e-scooters are not well investigated. However, using an e-scooter obviously requires less physical activity comparing to walking and cycling. Interestingly, social distancing during the pandemic was selected as the most important benefit of SFFESs. Recently, the COVID-19 virus hit Malaysia badly, and the number of new positive cases reached 4500 per day. This was the main concern of the current situation and people were seeking safe ways to go about their daily activities. 47% of survey participants indicated that they would not have car park issues by using SFFESs, and 45% believed that the environmental benefits (no pollution) of SFFESs were extremely important. In addition, “saving time” and “no traffic congestion” were indicated as extremely important benefits of SFFESs by 42% and 44% of participants, respectively.

The next series of questions were designed to ask respondents about their concerns about SFFESs, and what factors would prevent them from using this service, as presented in Figure 6. Safety was indicated as an extremely important concern of using SFFESs by 59% of respondents, moderately important by 26% of respondents, and not at all an important concern by only 4% of respondents. Surprisingly, the cost of riding SFFESs was selected as the second most important concern by 75% of respondents. In total, 53% and 22% of respondents indicated the “cost” as an extremely important and moderately important preventative factor, respectively. Due to the hot and humid tropical weather of Malaysia throughout the year, which is also interspersed with tropical rain showers, “adverse weather” is always a significant concern. Accordingly, almost 55% of respondents indicated the weather as an important preventative factor.

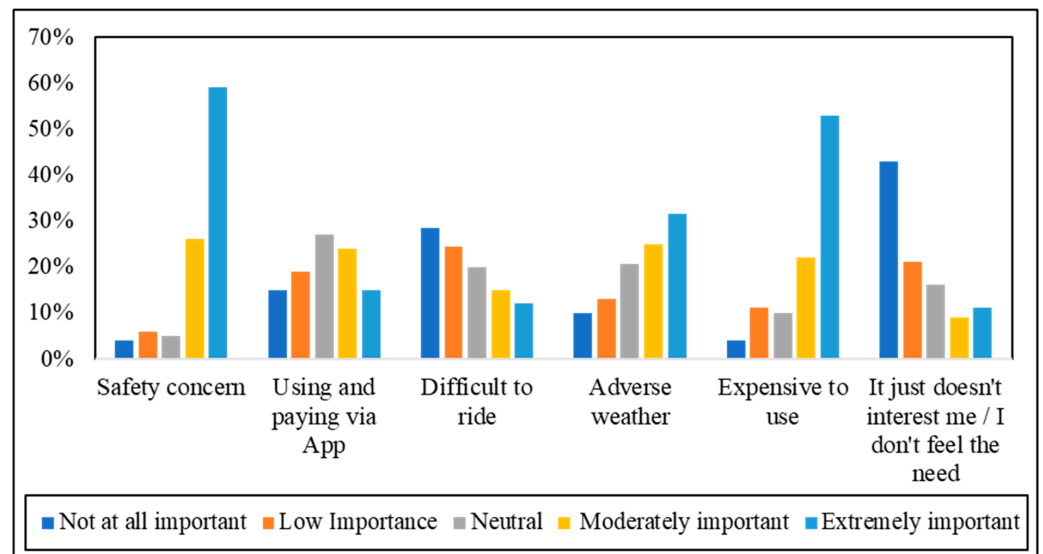


Figure 6. Participants’ opinions about what reasons would prevent them from using the SFESs.

As explained above, safety was indicated as the most important concern by almost 85% of the survey participants. Therefore, we decided to further explore this concern to gain better insights for policy making discussions and recommendations. Figure 7 illustrates SFES users’ perception of safety concerns based on their willingness to use the service in future. Respondents who would never ride e-scooters had the highest level of safety concern. Almost 40% of participants who belonged to this category specified that safety was an extremely important preventative factor to riding an e-scooter on campus, and 30% stated that it was moderately important.

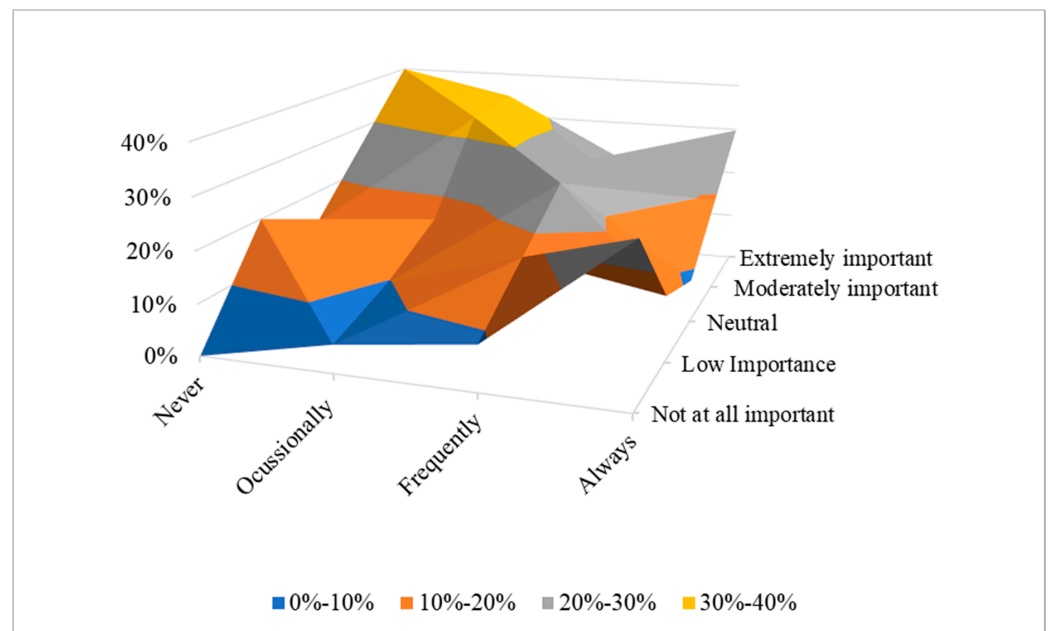


Figure 7. Safety concerns based on SFES usage categories.

In addition, over 80% of respondents who indicated safety as an extremely preventative concern also stated that they were extremely afraid of hitting somebody or being hit while riding an e-scooter. One of the chief causes of worry about accidents was the road features. Almost 60% of respondents who were extremely worried about safety indicated that separated scooter/bicycle lanes would strongly encourage them to ride an e-scooter.

In addition, almost 67% of them specified that no separated lanes for e-scooters would strongly discourage them from riding an e-scooter. The impact of other road features such as road connectivity, the quality of the surface and the availability of water and green spaces on their willingness to ride an e-scooter is shown in Figure 8.

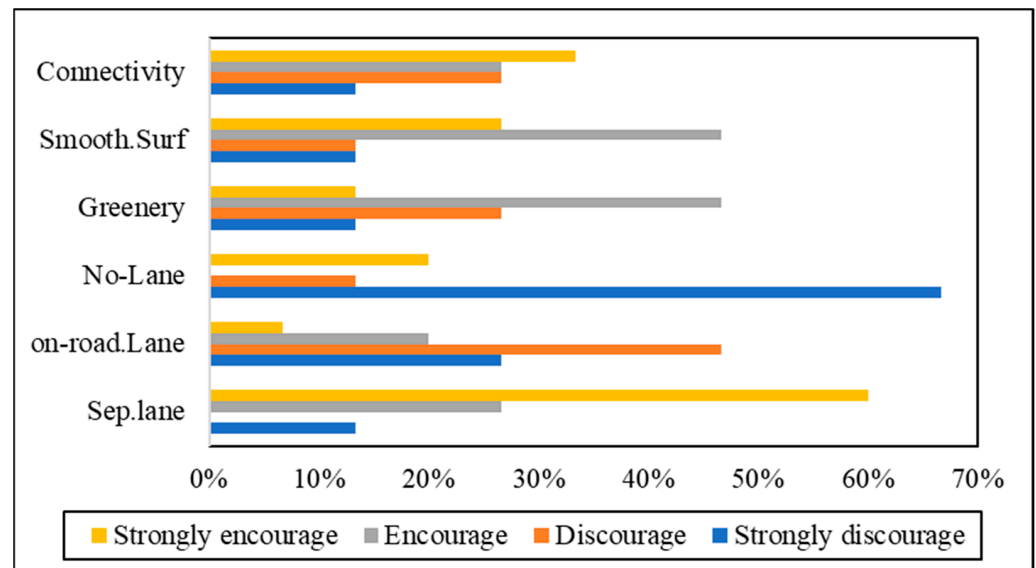


Figure 8. Impact of road features on the perceptions of respondents who believed safety was an extremely preventative factor for riding an e-scooter.

4.2. Policy Recommendation

In line with the intentions of the Malaysian government to develop green university campuses in the country, a number of universities in Malaysia have begun carrying out different green practices in an effort to improve sustainability. Accordingly, Malaysian academic centers, especially those at the higher education level, are dedicated to supporting the 40% reduction of carbon dioxide (CO₂) emissions vowed by the government at the 1992 Earth Summit in Rio [57]. Nevertheless, scholars such as [58] argue that, in Malaysian university management, practitioners and stakeholders are oblivious to green campus paradigms, which has caused most universities to ignore green practices. Currently, research on sustainability is initiated and socially certified by experts in higher education institutions [59]. However, there is still a lack of a proper method for interdisciplinary communication and cooperation among these sustainability practitioners to compile integrated data gleaned based on green indicators, which should be considered when achieving sustainability within Malaysian university campuses [60,61].

Nowadays, various sustainability practitioners in different areas of expertise work collaboratively to reach sustainability in the context of universities. However, interdisciplinary communication and collaboration is still absent among sustainability practitioners at higher education levels [62,63]. As [64] put, there is an urgent need for an interdisciplinary approach that is able to provide higher education institutions with a green campus paradigm toward accomplishing socio-economic and environmental sustainability. This is echoed by [65], who declared that there was insufficient harmonization and cooperation among practitioners from dissimilar domains that work jointly to obtain sustainability. The green campus concept aims to introduce engineering features including waste treatment, water treatment, and air pollution control, alongside personal aspects, such as promoting a laissez-faire outlook.

To develop a green campus, it is essential to assess the present data, information, and reports while focusing on enhancement. Generally, the aspects assessed in terms of green campus valuation instruments for higher education covers site and planning management, waste management, energy efficiency, sustainable transportation, water efficiency and

conservation, indoor environmental quality, material and resource management, green education, and green innovation. In this regard, electricity, waste generation, and transportation were chosen as targets considering their higher influences on CO₂ emission. Promoting active and novel modes of transportation can be an effective approach to reduce carbon emissions, as future transport will probably be dominated by electric vehicles (EVs). These vehicles offer several environmental benefits, which can lead to sustainability in urban transportation. More specifically, battery electric vehicles (BEVs) are gaining worldwide popularity. With their light weight, they could be well integrated into urban transport systems.

Electric scooters are emergent vehicles that could be used as an alternate transportation mode in campus and urban areas. These scooters have the potential to improve mobility and can be used in place of short car and ride-hail journeys. On the other hand, scooters have introduced some new challenges, which include safety, negative effects on disabled people, walkway clutter, etc. It is important for cities to evaluate the benefits that may be gained by using Shared Free-Floating Electric Scooter (SFFES) systems. SFFES services have the potential to introduce a number of environmental/social benefits, e.g., saving expenses and time (since they are generally faster than walking and even driving on crowded roads), lessening traffic blocking, enhancing multimodal transport connections, and decreasing the emissions of greenhouse gas (GHG). However, all these benefits are deeply dependent on adopted policies. For instance, based on our study results, most of the respondents specified the SFFES as an expensive transportation mode for campus usage. Making reliable decisions on this issue can be of great support to the expansion of e-scooter share programs in both campuses and cities.

4.3. Selection of Significant Variables through Unsupervised Clustering

Hierarchical clustering produced a dendrogram, which divided the 22 variables into 2 different clusters—13 variables in cluster one and 9 variables in cluster two. The variables in each cluster are shown in Figure 9.

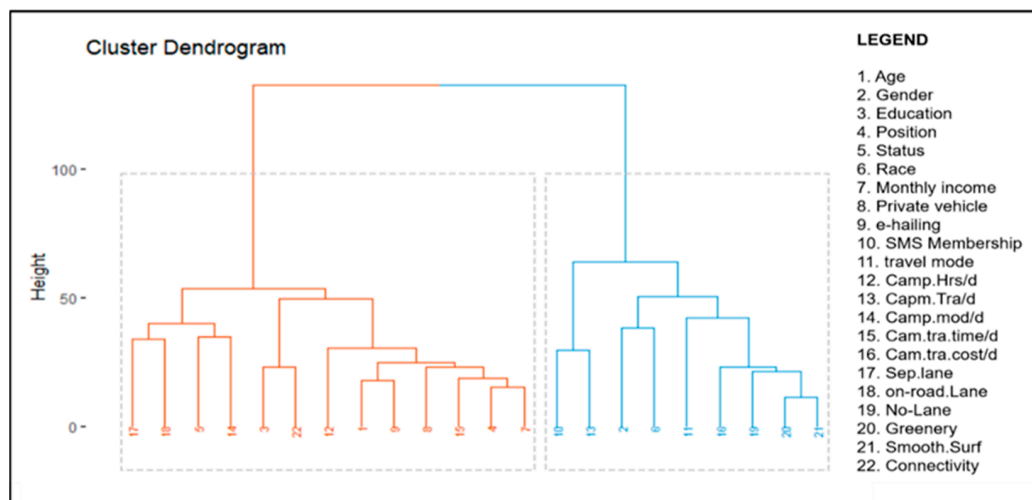


Figure 9. Cluster dendrogram of 22 variables.

The correlation between the variables was assessed using the dissimilarity matrix. The y-axis in the dendrogram in Figure 9 can be explained using the terms clade and leaves. The clusters were formed at a particular cluster cutoff value based on the number of clusters specified. As the analysis on determining optimal number of clusters regarding the dataset used in this study produced the result $k = 2$, the number of clusters was set as two. The specified number of clusters returned vectors containing features in each cluster. The lines showing the variables (number 1–22) are the leaves, whereas clusters 1 and 2 are clades 1 and 2 respectively. Leaves 17, 18, 5 and 14 are more similar to each other than they are

to other leaves in clade 1. Leaves 3 and 22 are more similar to each other than they are to other leaves in clade 1. Leaves 12, 1, 9, 8, 15, 4 and 7 are more similar to each other than they are to other leaves in clade 1. The x-axis in the dendrogram represents the clusters. The y-axis in the dendrogram represents the closeness of the leaves/variables. For example, leaves 4 and 7 were correlated before they joined 15, 8, and the following leaves together in one clade.

The distance between two clusters was measured using the linkage method. The complete linkage method used in this study displayed the distance between clusters 1 and 2 using the longest distance between two points in each cluster. The point refers to the line height in the dendrogram (Figure 9). The similarity between the features were assessed using the dissimilarity matrix index, whereas the important variables were determined using the line height. The heights of the lines in each leaf represent the importance score of the variables. In cluster 1, the most important features were Sep.lane, On-road.Lane, Status and Camp.mod/d with similar line heights. In cluster 2, the most important features were Gender, Race and Travel mode. To further assess the correlation between the independent variables, correlation analysis was performed. Figure 10 shows the correlation between the 22 independent variables.

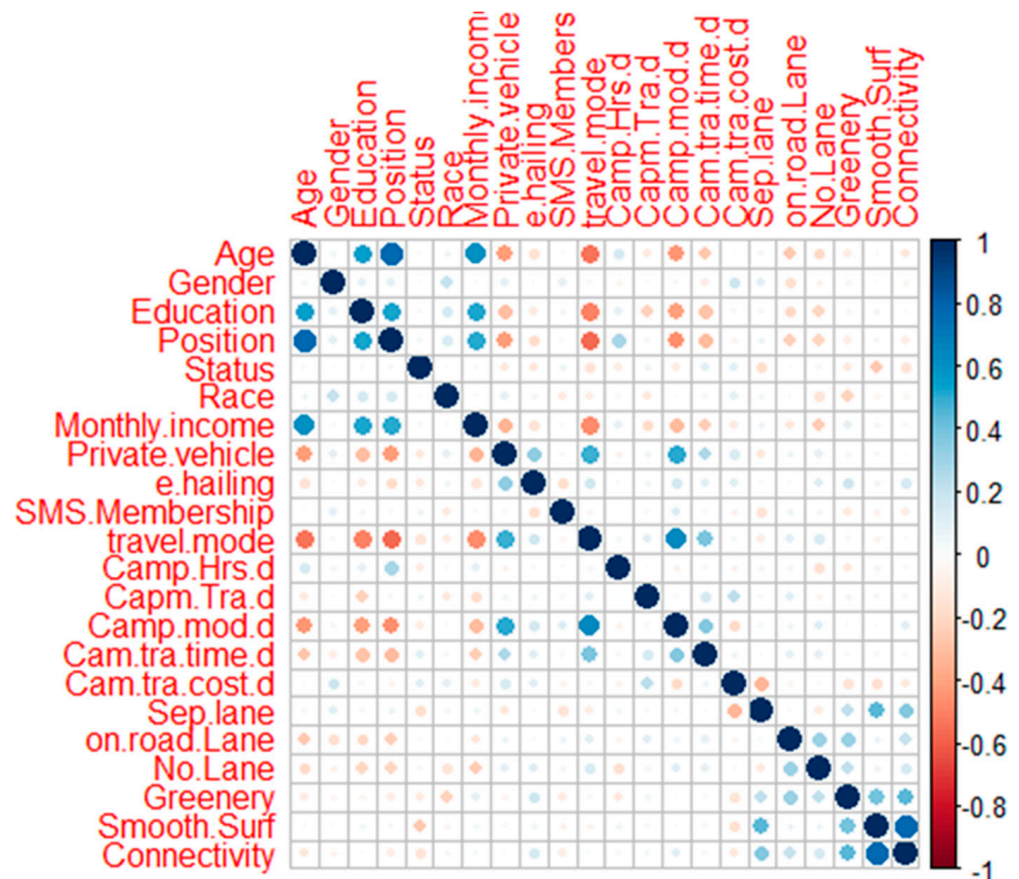


Figure 10. Correlation between the 22 independent variables.

The blue color represents positive correlation and the red color displays negative correlation. Based on the correlation analysis, two combinations are highly positively correlated, which are Position and Age and Connectivity and Smooth Surf. Moderately positively correlated combinations are Monthly Income and Age, and Camp.mod.d and travel.mode. Next, the lowly positively correlated combinations are Education and Age, Position and Education, Position and Monthly income, Monthly Income and Education, Private Vehicle and travel.mode, and Private Vehicle and Camp.mod.d.

4.4. Selection of Significant Variables Using Supervised Learning Models

The variables' importance rank of the 22 independent variables based on RF, DT and NB techniques are shown and compared in Figure 11. The present study takes advantage of various feature selection methods to pick only the important variables and design the prediction model according to selected variables. The core motive behind decreasing the number of variables (based on their level of importance and correlations) is to diminish the complexity and promote the applicability of our final model. Therefore, after implementing unsupervised clustering and identifying the correlation of the variables, we also compared the variables' importance based on three different tree-based supervised machine learning techniques. Table 3 presents the variable weights using outputs of RF, DT and NB. The mutually important variables were detected. For example, monthly income, age and private vehicle ownership were variables with high weights in all three methods.

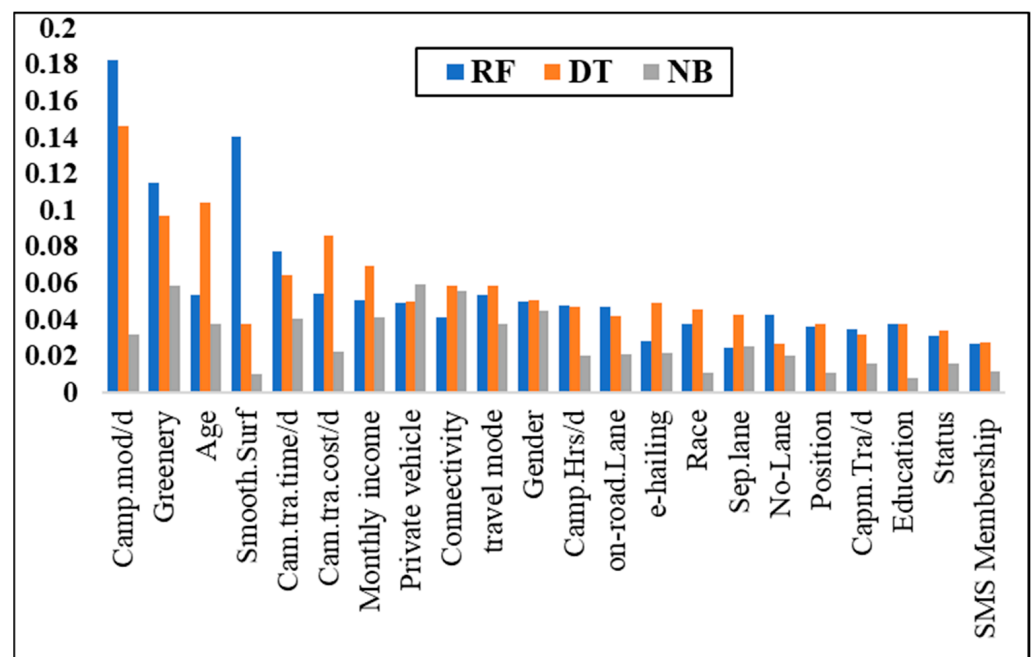


Figure 11. Importance score (weight) of variables based on three ML methods.

Furthermore, to make a more vivid conclusion from the three feature selection methods, the weight values of each variable were summarized and compared, as shown in Figure 12. Next, the amassed weight values were ranked from highest to lowest. According to Figure 12, there was a significant drop in weight values after the "Gender" variable. Therefore, we drew a line and deselected variables whose weights were below the line. The results of the selected most important variables based on three different ML techniques is summarized in Table 4.

Further random forest modelling was performed using these 11 variables. Moreover, all these variables have a threshold of MeanDecreaseGini higher than 30.

Table 3. Importance score (weight) of variables based on three ML methods.

No.	RF		DT		NB	
	Attribute	Weight	Attribute	Weight	Attribute	Weight
1	Camp.mod/d	0.1825	Camp.mod/d	0.14634	Private vehicle	0.059752
2	Smooth.Surf	0.1409	Age	0.10431	Greenery	0.058748
3	Greenery	0.1151	Greenery	0.09712	Connectivity	0.056134
4	Cam.tra.time/d	0.0777	Cam.tra.cost/d	0.08648	Gender	0.04504
5	Cam.tra.cost/d	0.0547	Monthly income	0.06964	Monthly income	0.041161
6	Travel mode	0.0538	Cam.tra.time/d	0.06434	Cam.tra.time/d	0.040235
7	Age	0.0534	Travel mode	0.0588	Travel mode	0.037877
8	Monthly income	0.0509	Connectivity	0.05861	Age	0.037555
9	Gender	0.0498	Gender	0.05055	Camp.mod/d	0.032029
10	Private vehicle	0.0490	Private vehicle	0.04992	Sep.lane	0.025078
11	Camp.Hrs/d	0.0477	e-hailing	0.04938	Cam.tra.cost/d	0.022726
12	on-road.Lane	0.0469	Camp.Hrs/d	0.04726	e-hailing	0.021732
13	No-Lane	0.0429	Race	0.0454	on-road.Lane	0.021263
14	Connectivity	0.0415	Sep.lane	0.04247	No-Lane	0.020533
15	Race	0.0376	on-road.Lane	0.04189	Camp.Hrs/d	0.019974
16	Education	0.0374	Position	0.03798	Capm.Tra/d	0.016223
17	Position	0.0366	Smooth.Surf	0.03751	Status	0.015691
18	Capm.Tra/d	0.0351	Education	0.0374	SMS Membership	0.011859
19	Status	0.0308	Status	0.03386	Position	0.010624
20	e-hailing	0.0280	Capm.Tra/d	0.03175	Race	0.010495
21	SMS membership	0.0268	SMS Membership	0.02767	Smooth.Surf	0.010309
22	Sep.lane	0.0248	No-Lane	0.02676	Education	0.0082

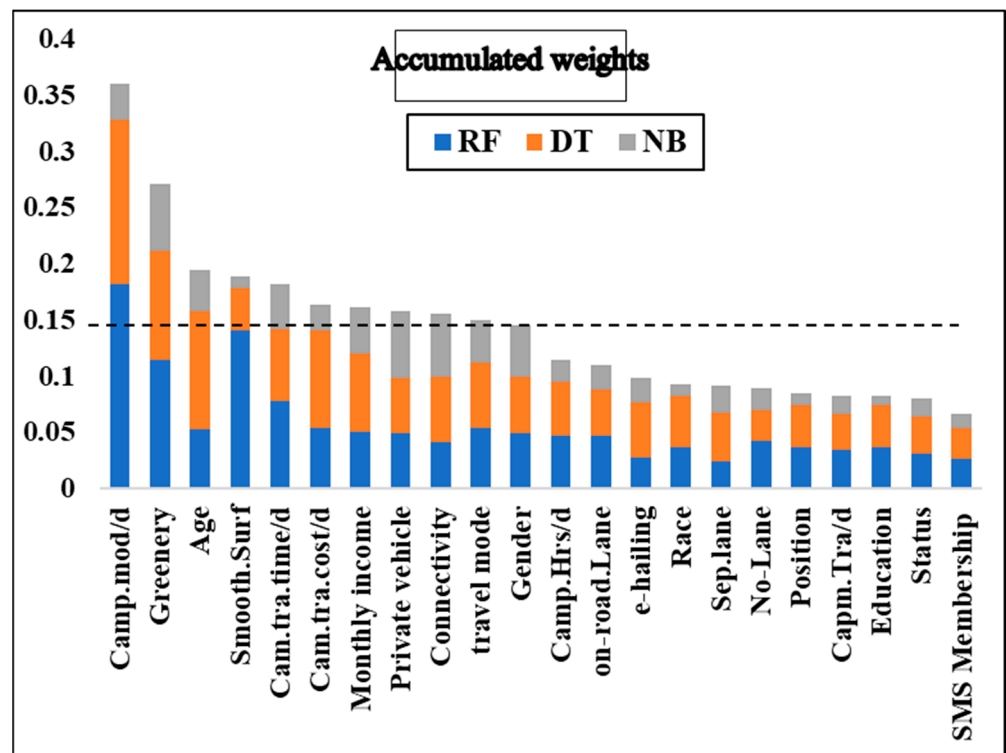


Figure 12. Accumulated weights of variables.

Table 4. Importance of 11 selected variables based on feature selection criteria.

No.	Attribute	Accumulated Weight	Mean Decrease Gini
1	Camp.mod/d	0.360867959	72.26206
2	Greenery	0.270941347	62.26634
3	Age	0.195234017	61.92460
4	Smooth.Surf	0.188729931	60.28623
5	Cam.tra.time/d	0.182241979	59.64285
6	Cam.tra.cost/d	0.16393153	57.96135
7	Monthly income	0.161725573	57.71634
8	Private vehicle	0.158708056	53.55493
9	Connectivity	0.156257276	51.93130
10	Travel mode	0.150511383	44.97282
11	Gender	0.145347998	44.94371

4.5. Model Assessment and Evaluation

Having reduced the number of variables by a comprehensive feature selection method (through both unsupervised clustering and supervised models), the random forest algorithms were conducted using eleven selected variables. The model performance of random forest is reported as below:

Call:

Number of trees: 500

No. of variables tried at each split: 3

Mean of squared residuals: 0.07049505

% Var explained: 93.02

The default *ntree* used was 500 and *mtry* was 3. The accuracy was 93.02% and the Mean of squared residuals was 0.07049505. The error vs number of tree graphs in Figure 13 shows that the error rate remained constant from 390 to 470. Model assessment was repeated nine times using a different number of trees from 390 to 470, and the results are presented in Table 5.

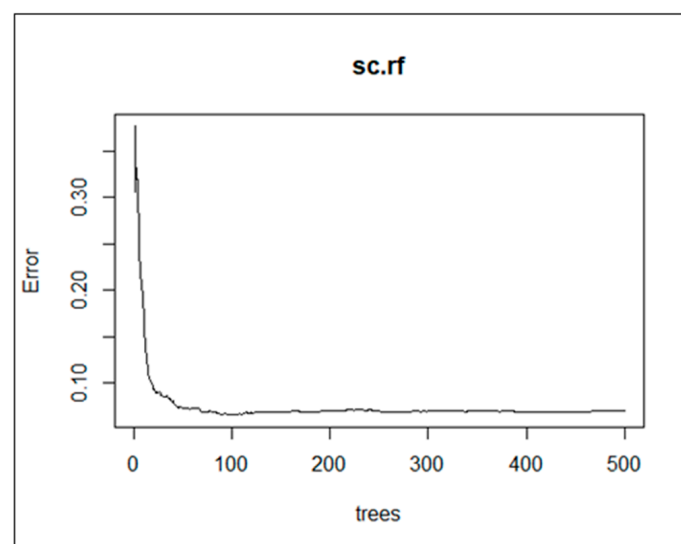


Figure 13. The error vs number of tree graphs for 11 important features.

Table 5. Random forest model.

No	Number of Trees	Accuracy (%)
1	390	93.42
2	400	93.26
3	410	93.28
4	420	93.19
5	430	93.29
6	440	93.51
7	450	93.14
8	460	93.15
9	470	93.28

The best *ntree* was 440 as shown in Table 6, since it produced the highest accuracy compared to other values. The *ntree* = 440 was used to assess the test error and OOB error, as shown in Figure 14.

Table 6. Model assessment for decision tree, random forest and Naïve Bayes.

Model	Algorithm	Accuracy (%)		Precision		Recall		F1 Score	
		11 Variable	22 Variable	11 Variable	22 Variable	11 Variable	22 Variable	11 Variable	22 Variable
Decision tree	<i>rpart</i> from "caret"	54.13	57.130	0.29	0.318	0.38	0.4000	0.32	0.325
Random Forest	<i>rf</i> from "caret"	93.51	99.49	0.85	0.890	0.82	0.850	0.72	0.760
Naïve Bayes	<i>nb</i> from "e1071" package	61.00	64.50	0.51	0.530	0.45	0.480	0.52	0.540

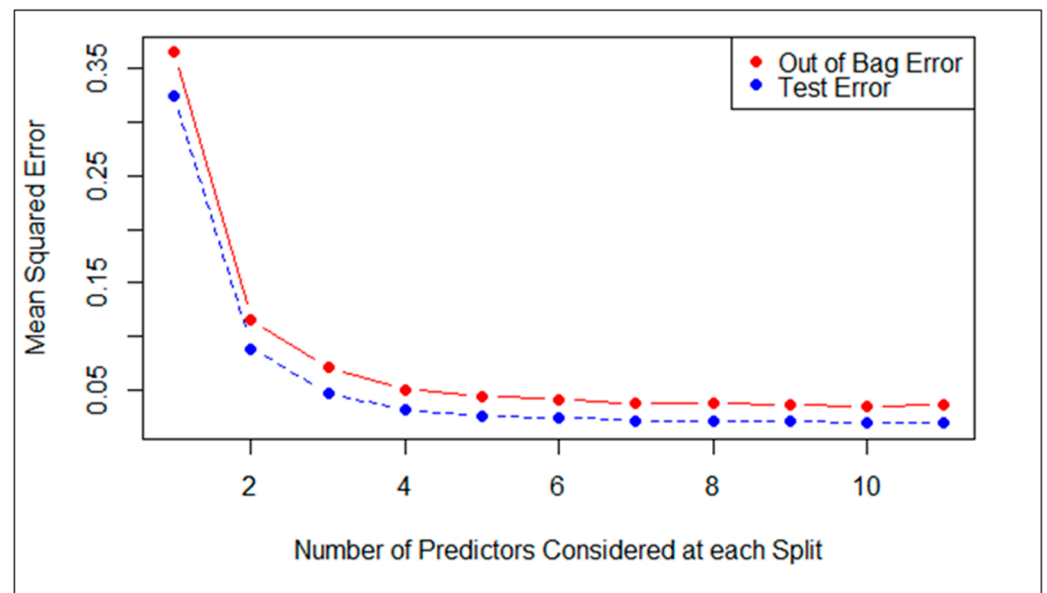


Figure 14. Test error and out-of-bag (OOB) error rate of the predicted model.

The red line represents the out-of-bag error estimates, and the blue line represents the error calculated on test set. Both curves are relatively smooth, and the error estimates are also correlated. The error inclines are reduced at around *mtry* = 3. Hence, the final model with the 11 most important variables produced an accuracy of 93.51%, with *ntree* = 440 and *mtry* = 3. The model performance comparison among the random forest, decision tree and Naïve Bayes methods are shown in Table 6 for both models with 22 variables and 11 variables.

4.6. Simulation-Based Optimization Analysis

To obtain deeper insights into the factors influencing SFFES usage, optimization analysis was carried out based on four different scenarios: a group of respondents who are: (1) most likely to “always” use SFFESs, (2) most likely to “frequently” use SFFESs, (3) most likely to “occasionally” use SFFESs, and (4) less likely to, or “never”, use SFFESs. The scenarios were based on the target variables’ response categories, as described in Table 1. The simulation-based optimization analysis was conducted on 11 significant variables (as described in the feature selection section) using RapidMiner Studio Educational Software version 9.8.001. All the figures in this section are outputs of the RapidMiner Software. The optimization was carried out and determined the best input factors to fit with our targets under the specified constraints. Additionally, the simulation-based sensitivity analysis was considered appropriate for evaluating and responding “What if” questions. For example, what if our target group is male youngsters who are between 18 to 23 years old and who use public transportation as their mode of transport on the campus (how frequently they will use SFFESs)? Table 7 presents the optimized value of attributes based on four scenarios.

Table 7. Optimized value of attributes based on four scenarios.

Attribute	Always	Frequently	Occasionally	Never
Gender	Female	Female	Male	Male
Age	18 to 29	30 to 44	45 to 60	45 to 60
Monthly income	Between RM 4000 and RM 6000	Between RM 6000 and RM 12,000	Between RM 2000 RM 4000	Between RM 6000 and RM 12,000
Travel mode	Walking/cycling	Public transportation	Private car	Private car
Private vehicle	No	Yes	Yes	Yes
Camp.mod/d	Walking/cycling	E-hailing	Public Transport	Private car
Cam.tra.cost/d	Between RM 5 and RM15	Between RM 15 and RM 25	Less than RM 5	Less than RM 5
Cam.tra.time/d	20 to 30 min	Less than 10 min	10 to 20 min	Less than 10 min
Greenery	Encourage	Strongly encourage	Strongly discourage	Encourage
Smooth.Surf	Encourage	Discourage	Encourage	Encourage
Connectivity	Encourage	Encourage	Discourage	Encourage

In the first scenario, the simulation model was adjusted to optimize the target variables on respondents who are most likely to always use SFFESs. According to the results, females between 18 and 29 years old with a monthly income between RM 4000 and RM 6000 (which is a higher-than-average income in Malaysia), whose primary mode of transport is walking or cycling, are the most likely to change their mode of transport to SFFESs. This group of respondents does not own a private vehicle and they spend RM 5 to RM 15 for their travels around the campus per day.

According to Figure 15, 95% of respondents described above will always use SFFESs as their main mode of transport on the campus, 3% will use SFFESs occasionally, 1.5% will never use it, and less than 1% will use it frequently. In addition, gender, age, and cost of travel per day are the most important factors affecting SFFESs choice and usage. The simulation model was adjusted to optimize the attributes based on the second scenario and determine the characterization of the SFFES service’s frequent users. Frequent usage of the SFFES service has been defined as usage between two and five times per week, or replacing at least half of the user’s current mode of transport with the SFFES service. According to Table 7, most of the frequent users of SFFESs will be women, as in the previous scenario. However, frequent users are most likely to be older users (30 to 40 years old) with a higher monthly income. While they most likely own private vehicles, they mostly use public transportation for arriving on campus and use e-hailing services to travel around the campus. According to Figure 16, 77% of described students/staff are willing to use the SFFES service frequently. In addition, road features such as connectivity and quality of road surface can strongly impact their usage. Travel mode and travel costs are other important factors for this group, according to Figure 16.

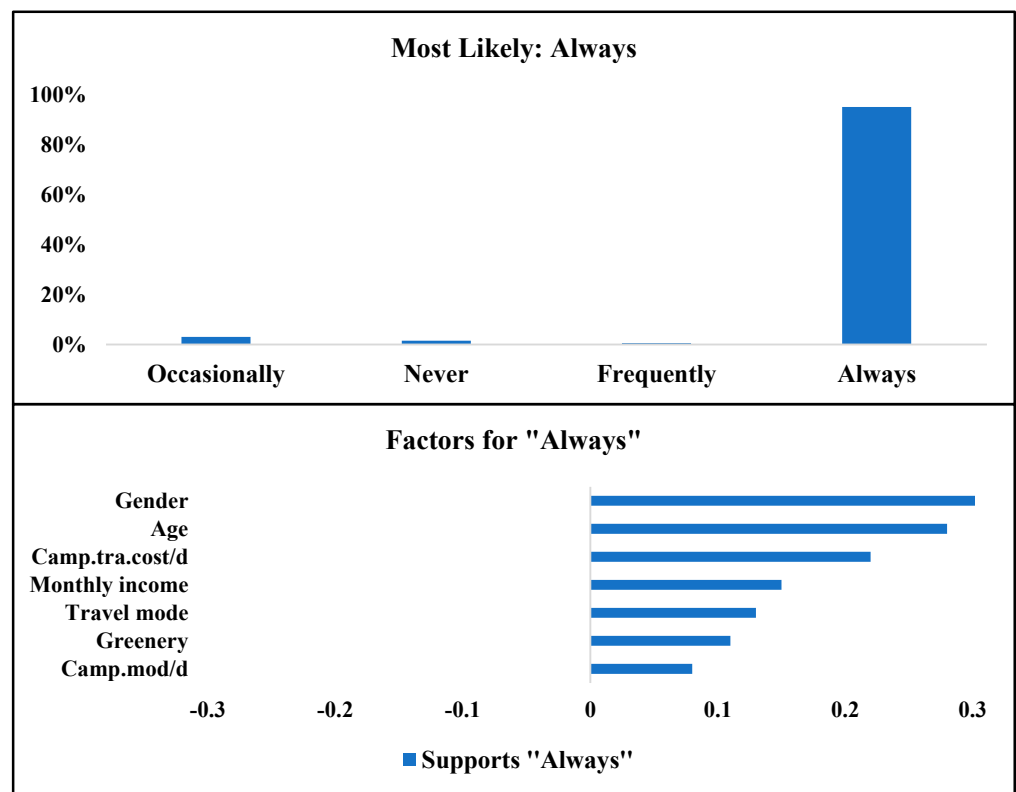


Figure 15. Optimization results and importance of variables based on the first scenario: Always use SFFESs.

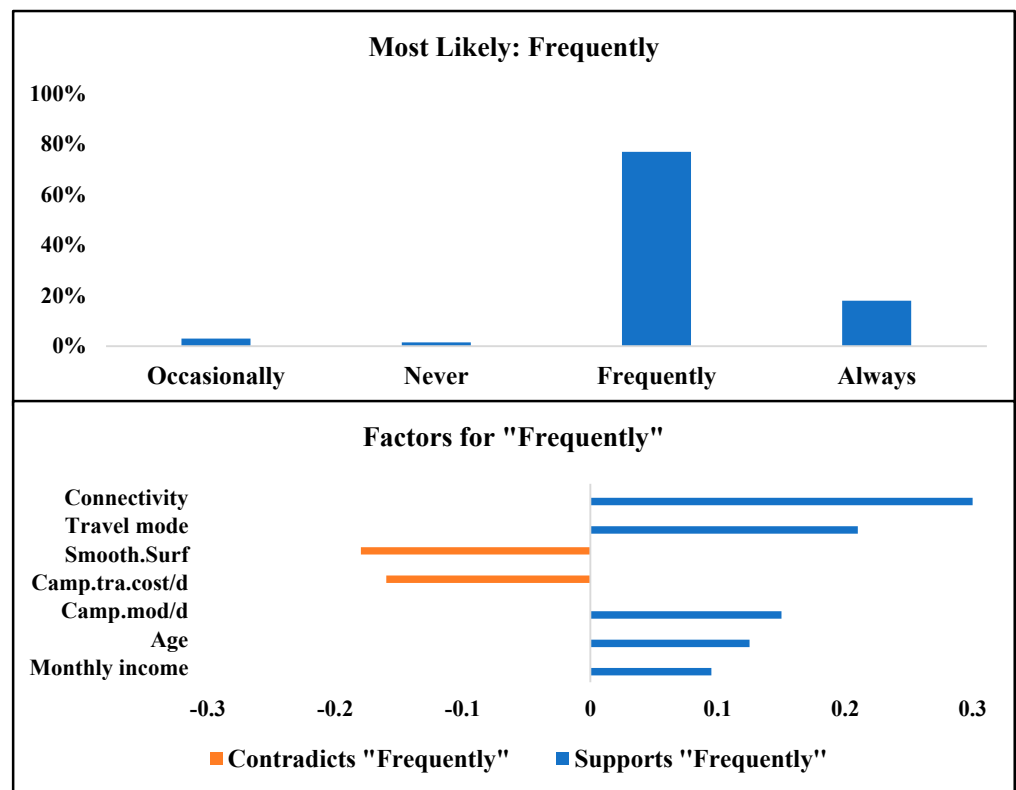


Figure 16. Optimization results and importance of variables based on the second scenario: Frequently use SFFESs.

In the third scenario, the simulation model optimized target variables on the group of users who will most likely use SFFESs occasionally (less than three times per week). Interestingly, men between 45 and 60 years old with an average monthly salary (RM 2000 to RM 4000 is considered an average monthly income in Malaysia) are most likely to use SFFESs occasionally. In addition, they own private vehicles and mostly use public transportation for their daily travels around the campus. According to Figure 17, 82% of users who are described in the third scenario will use SFFES services occasionally or less than three times per week. Moreover, travel mode, age and daily travel time are important factors which support their SFFES mode choice.

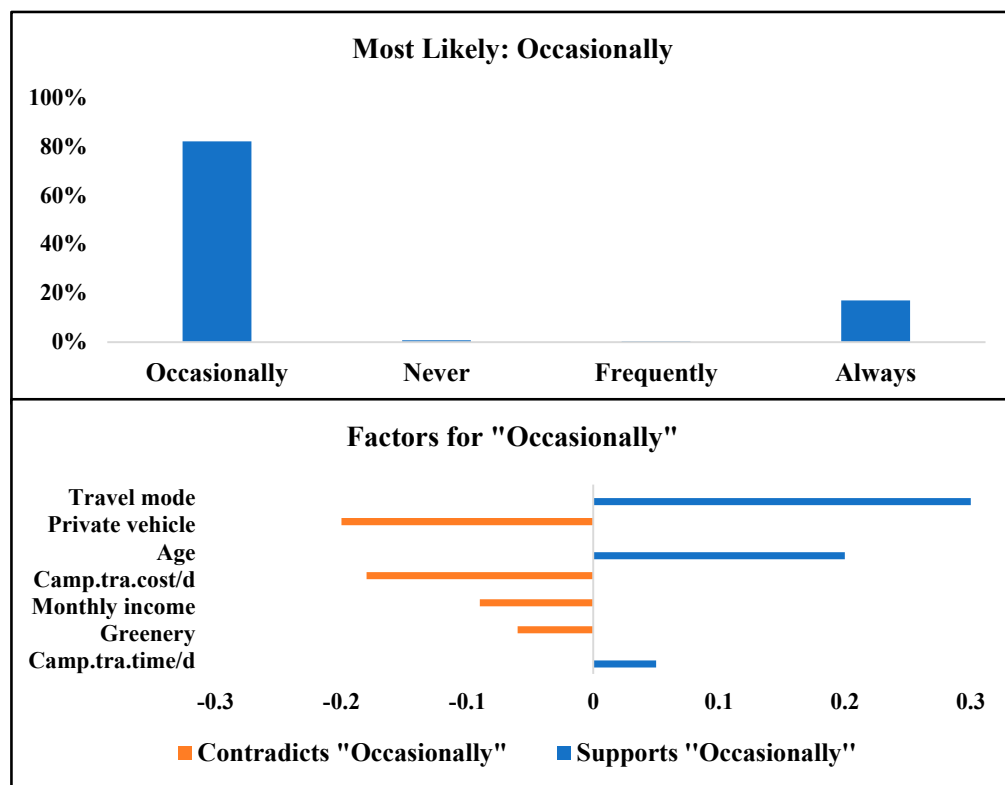


Figure 17. Optimization results and importance of variables based on the second scenario: Occasionally use SFFESs.

Respondents who are not interested in SFFESs and will never use the service were our target in the fourth scenario. According to the last column of Table 7, the sociodemographic characterization of respondents in this scenario is almost the same as the third scenario (users who will use SFFESs occasionally), with the difference being that their monthly income is much higher. In addition, their average daily travel time is significantly shorter, and they prefer to use their own car. As shown in Figure 18, 89% of users described in the fourth scenario are most likely to never use SFFESs. Moreover, road features such as green roads and smooth surfaces are the most important factors which are in contrast with the "Never" usage scenario. In other words, road features are significantly important factors that may encourage them to consider SFFES services for their future travels around the campus (as shown in Figure 18).

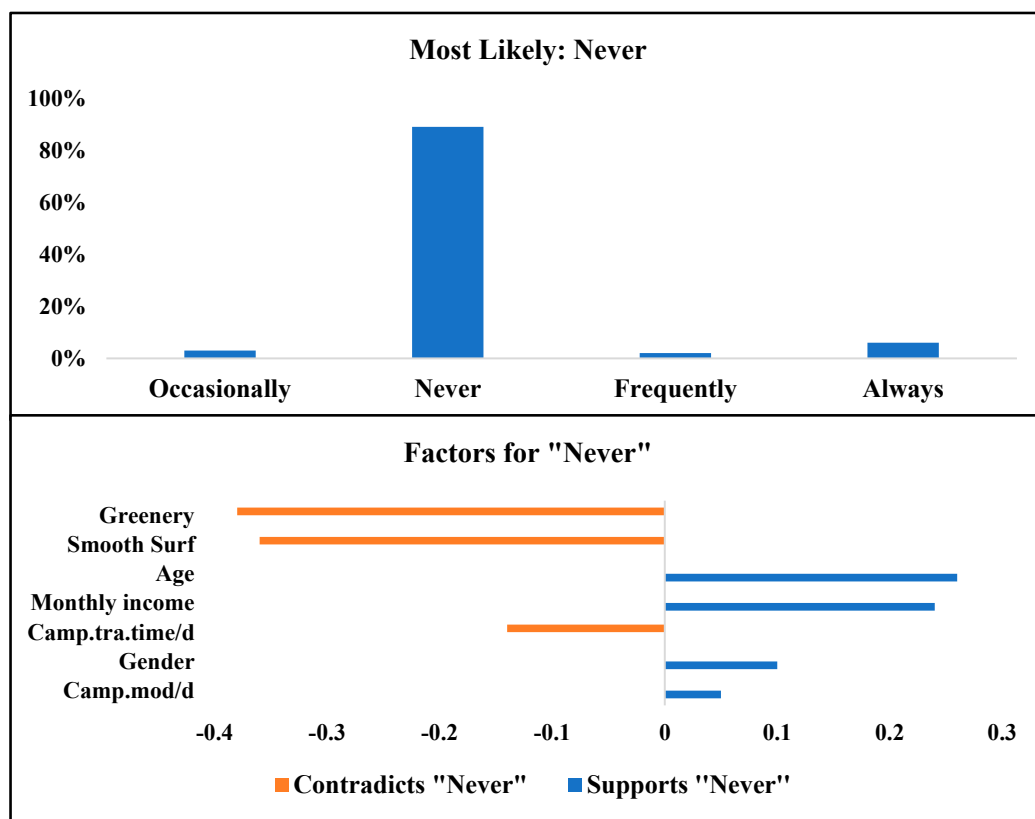


Figure 18. Optimization results and importance of variables based on the second scenario: Never use SFFESs.

5. Discussion

This study is designed to deeply explore the attitudes and perceptions of students and staff towards SFFES usage on campus. Various attributes were considered for this propose, such as the sociodemographic information of respondents, characterization of trips, road features, concerns/barriers, and benefits of riding SFFESs. In addition, this study is one of the first to predict the likelihood of usage frequency of SFFESs by employing various machine learning techniques and the first study on SFFESs in Malaysia. Previous studies have been mostly conducted in the US, China, and recently European cities. Moreover, for the first time we have employed different feature selection methods and machine learning algorithms to deeply evaluate the weight of important factors that affect the mode choice and usage of SFFESs between university students and staff. The campus of University of Malaya (UM) was selected for conducting this study because:

- Shared micromobility is new in Malaysia, and most people have limited knowledge about it. The university community is a natural laboratory to test new mobility services.
- The shared e-scooter companies such as BEAM, TRYKE and Myscooter are very interested in providing their services to university campuses in this initial stage.
- UM is the biggest university in Malaysia, with more than 30,000 students and staff. In addition, more than 5000 international students and staff are on UM campus of different races, ethics, nations and generations. The diversity of the population fits the study requirements well.

A comprehensive feature selection was conducted before developing machine learning predictive models. The main propose of this step was to accurately recognize the significant factors and importance by adopting supervised and unsupervised machine learning techniques. In addition, it decreased the complication of the final model by decreasing the number of variables based on their significance. Although decreasing the number of variables may reduce the accuracy of the final model, this reduction can be minimized

by adopting proper and accurate feature selection techniques. In other words, adopting accurate feature selection methods will promote the complexity and practicality of the final model, while the accuracy remains adequately high. In this study, the initial models were developed using all 22 variables. After implementing feature selection methods, the number of variables reduced to 11.

According to the feature selection results, daily travel mode inside the campus (Camp.mod/d) was the most effective factor in determining SFFES usage frequency. Other travel characterizations, such as daily travel cost and time/duration, were among the most influential factors. Sociodemographic attributes such as age, gender, monthly income and private vehicle ownership, played significantly important roles in mode choice and SFFES usage, as also proved by previous studies [33,66]. In addition, based on the outputs of all three supervised feature selection models, road features such as greenery and the connectivity of roads influenced the mode choice significantly. The initial RF model (with 22 variables) outperformed DT and NB models with 99.45% accuracy. Therefore, we selected the RF model for further analysis and developing the final model using the 11 most important variables. As expected, reducing the number of variables caused a reduction in accuracy by 6%. However, the authors believe that the final model is a much more valuable model with acceptable accuracy and less complexity.

To unpack and shed light on the attitudes of the survey participants towards SFFES usage, a simulation-based optimization was developed. Interesting results have been gained which could be useful for future works, recommendations and policy-making. Four optimization scenarios were defined based on the four categories of possible SFFES users: always, frequently, occasionally and never. According to the optimization results, there was a strong relationship between gender and the frequency of usage of SFFESs. Surprisingly, respondents who were more likely to ride e-scooters always and frequently were mostly young to middle-aged females. This result is in contrast with previous studies [15,67] and further exploration is needed to discover the reasons for this gender gap in SFFES usage. However, this result may be biased by our survey participants' characterization, who were mostly young and highly educated.

Indeed, many interesting facts can be unveiled by adopting simulation-based optimization analysis. According to Table 6, respondents who would change their travel mode to SFFESs were mostly daily cyclists or pedestrians. The same result was observed by previous studies [15,33,68]. On the other hand, respondents who used their own private vehicle for daily trips were not interested in riding an e-scooter. These two facts can be considered as significant disadvantages of SFFES services. Undoubtedly, walking and cycling are more desirable and sustainable modes of transportation in several different aspects. Walking and cycling are healthier modes, since they require much more physical activity [69]. Moreover, while walking and cycling are the greenest possible modes of transport, the environmental impact of e-scooters is still not well-investigated [70].

Strength, Limitations and Next Steps

Before indicating the limitations, the authors would like to mention the significant strengths of this study. To the authors' knowledge, this is the first study of SFFESs on a university campus. A large number of students and staff with various sociodemographic backgrounds and undertaking different types of activities on the campus helped to shed some light on the future of SFFESs launches on other university campuses and even urban areas. Furthermore, this was the first study on SFFESs undertaken in Malaysia and one the first to employ various machine learning algorithms to predict the use frequency of SFFESs. There are also a number of limitations. One of the key limitations of this study was sample size. We forwarded the Google Form (the survey) to more than 30,000 university students and staff, and only 1.7% responded completely. The number of respondents was limited, and there must also be principal differences between respondents and non-respondents. In addition, the method of survey distribution and focus group was limited to academic and highly educated people. Undoubtedly, further studies should consider larger sample

sizes which are more random and representative of potential SFFES riders. Moreover, we have not provided some specific scenarios for using SFFESs on the campus, such as estimated travel time, costs, proposals and external factors like weather. Therefore, the answers to some questions were based on the experience of respondents, which would influence the results.

Future studies should consider larger sample sizes to predict a better model with higher accuracy, which also represents all SFFES users in Malaysia. Moreover, future studies should incorporate the available information from SFFES companies, such as travel distance, travel time and proposals for travel. In this study, we have only focused on tree-based machine learning algorithms (RF, DT and NB) for predicting SFFES usage frequency. We propose that future studies should consider other types of machine learning techniques, such as neural networks and support vector machines, to clarify which technique has the best performance. Finally, off-campus and on-campus students have different requirements, and in turn, travel behaviors. Future studies can consider these differences.

6. Conclusions

This study predicts SFFES use on a university campus using supervised and unsupervised machine learning techniques. A comprehensive feature selection analysis was conducted using k-means and hierarchical clustering, decision tree, random forest and Naïve Bayes techniques. The 11 most important attributes were identified, including daily travel modes around the campus, the presence of green spaces and water, age, quality of the road surface, daily travel time and cost around the campus, monthly income, private vehicle ownership, connectivity between roads, modes of transport to/from campus, and gender.

The random forest algorithm was developed to predict the usage frequency of SFFES using the identified important attributes. Simulation-based sensitivity analysis was conducted to gain deeper insights into the characterization and specification of SFFES users. Young females between 18 and 29 years old with an average monthly income were the most likely to always use SFFESs for their travels on campus. Males between 45 and 60 years with a high monthly salary were less likely to use SFFESs. The safety concerns and cost of renting e-scooters were the most important discouragement factors, while road features and suitable infrastructure, such as green spaces and separated lanes for scooters, were the most important encouragement factors. In addition, social distancing during the pandemic and no parking issues were the most considerable benefits of riding e-scooters from the respondents' perspective.

The responsibility of the service providers and authorities is to provide all residents (especially people with limited transportation access) with accessible, equitable, safe, affordable, and sustainable transportation options. SFFES services are capable of helping to fill transportation gaps through providing an efficient, affordable alternative to cars for urban journeys. Scooters can have several benefits such as health, safety, and congestion opening, as well as some social/environmental equity benefits. To make an effective decision regarding whether and how SFFESs should be implemented in transportation systems of future cities, decision makers must first determine the definite role of these vehicles in the city. This can be determined through finding out the involved actors' visions of the future urban transport. As a result, to guarantee sustainable mobility, there is a need for not only technology and investment, but also fundamental research into related issues.

Author Contributions: S.M.H.M., conceptualization, methodology, software, formal analysis, investigation, resources, writing—original draft, supervision; Z.M., investigation, writing—review and editing, supervision; D.J.A., conceptualization, software, writing—review and editing, supervision; M.A., formal analysis, writing—review and editing; M.D.G., conceptualization, software, investigation; Y.C.W., conceptualization, resources, supervision; D.V.U., writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available from the corresponding author upon reasonable request.

Acknowledgments: We would like to acknowledge all the experts and staff in the BEAM scooter company and the University of Malaya for providing data and information. In particular, the authors would like to acknowledge the Centre for Transportation Research (CTR), Faculty of Engineering, and University of Malaya for providing research facilities.

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

References

1. Kalda, K.; Pizzagalli, S.-L.; Soe, R.-M.; Sell, R.; Bellone, M. Language of Driving for Autonomous Vehicles. *Appl. Sci.* **2022**, *12*, 5406. [CrossRef]
2. Shaheen, S.; Cohen, A.; Chan, N.; Bansal, A. Sharing strategies: Carsharing, shared micromobility (bikesharing and scooter sharing), transportation network companies, microtransit, and other innovative mobility modes. In *Transportation, Land Use, and Environmental Planning*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 237–262.
3. Fitt, H.; Curl, A. The early days of shared micromobility: A social practices approach. *J. Transp. Geogr.* **2020**, *86*, 102779. [CrossRef]
4. Kou, Z.; Wang, X.; Chiu, S.F.A.; Cai, H. Quantifying greenhouse gas emissions reduction from bike share systems: A model considering real-world trips and transportation mode choice patterns. *Resour. Conserv. Recycl.* **2020**, *153*, 104534. [CrossRef]
5. Li, W.; Kamargianni, M. Providing quantified evidence to policy makers for promoting bike-sharing in heavily air-polluted cities: A mode choice model and policy simulation for Taiyuan-China. *Transp. Res. Part A Policy Pract.* **2018**, *111*, 277–291. [CrossRef]
6. Lazarus, J.; Pourquier, J.C.; Feng, F.; Hammel, H.; Shaheen, S. Micromobility evolution and expansion: Understanding how docked and dockless bikesharing models complement and compete—A case study of San Francisco. *J. Transp. Geogr.* **2020**, *84*, 102620. [CrossRef]
7. McKinsey & Co. Sizing the Micro Mobility Market | McKinsey. McKinsey & Co. 2021. Available online: <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/micromobilitys-15000-mile-checkup> (accessed on 7 February 2021).
8. Berg Insight. The Bike and Scootersharing Telematics Market. 2020; pp. 2–5. Available online: <http://www.bergsinsight.com/ReportPDF/ProductSheet/bi-micromobilitytelematics2-ps.pdf> (accessed on 1 March 2020).
9. Tuncer, S.; Laurier, E.; Brown, B.; Licoppe, C. Notes on the practices and appearances of e-scooter users in public space. *J. Transp. Geogr.* **2020**, *85*, 102702. [CrossRef]
10. Sgarbossa, F.; Peron, M.; Fragapane, G. Cloud material handling systems: Conceptual model and cloud-based scheduling of handling activities. *Int. Ser. Oper. Res. Manag. Sci.* **2020**, *289*, 87–101. [CrossRef]
11. Lolli, F.; Coruzzolo, A.M.; Peron, M.; Sgarbossa, F. Age-based preventive maintenance with multiple printing options. *Int. J. Prod. Econ.* **2022**, *243*, 108339. [CrossRef]
12. Mont, O.; Palgan, Y.V.; Bradley, K.; Zvolzka, L. A decade of the sharing economy: Concepts, users, business and governance perspectives. *J. Clean. Prod.* **2020**, *269*, 122215. [CrossRef]
13. Nguyen-Phuoc, D.Q.; Amoh-Gyimah, R.; Tran, A.T.P.; Phan, C.T. Mode choice among university students to school in Danang, Vietnam. *Travel Behav. Soc.* **2018**, *13*, 1–10. [CrossRef]
14. Rotaris, L.; Danielis, R.; Maltese, I. Carsharing use by college students: The case of Milan and Rome. *Transp. Res. Part A Policy Pract.* **2019**, *120*, 239–251. [CrossRef]
15. Sanders, R.L.; Branion-Calles, M.; Nelson, T.A. To scoot or not to scoot: Findings from a recent survey about the benefits and barriers of using E-scooters for riders and non-riders. *Transp. Res. Part A Policy Pract.* **2020**, *139*, 217–227. [CrossRef]
16. Stylianou, K.; Dimitriou, L.; Abdel-Aty, M. Big data and road safety: A comprehensive review. In *Mobility Patterns, Big Data and Transport Analytics*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 297–343.
17. Yang, H.; Song, K.; Zhou, J. Automated Recognition Model of Geomechanical Information Based on Operational Data of Tunneling Boring Machines. *Rock Mech. Rock Eng.* **2022**, *55*, 1499–1516. [CrossRef]
18. Yang, H.; Wang, Z.; Song, K. A new hybrid grey wolf optimizer-feature weighted-multiple kernel-support vector regression technique to predict TBM performance. *Eng. Comput.* **2020**, *38*, 2469–2485. [CrossRef]
19. Du, K.; Liu, M.; Zhou, J.; Khandelwal, M. Investigating the slurry fluidity and strength characteristics of cemented backfill and strength prediction models by developing hybrid GA-SVR and PSO-SVR. *Min. Metall. Explor.* **2022**, *39*, 433–452. [CrossRef]
20. Zhou, J.; Qiu, Y.; Khandelwal, M.; Zhu, S.; Zhang, X. Developing a hybrid model of Jaya algorithm-based extreme gradient boosting machine to estimate blast-induced ground vibrations. *Int. J. Rock Mech. Min. Sci.* **2021**, *145*, 104856. [CrossRef]
21. Zhou, J.; Li, X.; Mitri, H.S. Classification of rockburst in underground projects: Comparison of ten supervised learning methods. *J. Comput. Civ. Eng.* **2016**, *30*, 4016003. [CrossRef]
22. Parsajoo, M.; Armaghani, D.J.; Mohammed, A.S.; Khari, M.; Jahandari, S. Tensile strength prediction of rock material using non-destructive tests: A comparative intelligent study. *Transp. Geotech.* **2021**, *31*, 100652. [CrossRef]
23. Hasanipanah, M.; Monjezi, M.; Shahnezar, A.; Armaghani, D.J.; Farazmand, A. Feasibility of indirect determination of blast induced ground vibration based on support vector machine. *Measurement* **2015**, *75*, 289–297. [CrossRef]

24. Toch, E.; Lerner, B.; Ben-Zion, E.; Ben-Gal, I. Analyzing large-scale human mobility data: A survey of machine learning methods and applications. *Knowl. Inf. Syst.* **2019**, *58*, 501–523. [CrossRef]
25. Xu, C.; Ji, J.; Liu, P. The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transp. Res. Part C Emerg. Technol.* **2018**, *95*, 47–60. [CrossRef]
26. Gao, X.; Lee, G.M. Moment-based rental prediction for bicycle-sharing transportation systems using a hybrid genetic algorithm and machine learning. *Comput. Ind. Eng.* **2019**, *128*, 60–69. [CrossRef]
27. Aghaabbasi, M.; Shekari, Z.A.; Shah, M.Z.; Olakunle, O.; Armaghani, D.J.; Moeinaddini, M. Predicting the use frequency of ride-sourcing by off-campus university students through random forest and Bayesian network techniques. *Transp. Res. Part A Policy Pract.* **2020**, *136*, 262–281. [CrossRef]
28. Čuš-Babič, N.; de Oliveira, S.F.G.; Tibaut, A. Interoperability of Infrastructure and Transportation Information Models: A Public Transport Case Study. *Appl. Sci.* **2022**, *12*, 6234. [CrossRef]
29. Bokolo, A.J. Green campus paradigms for sustainability attainment in higher education institutions—A comparative study. *J. Sci. Technol. Policy Manag.* **2020**, *12*, 117–148. [CrossRef]
30. Zakaria, R.; Alqaifi, G.; Rahim, A.; Hamid, A.R.A.; Mansur, S.A.; Resang, A.; Zen, I.S.; Bandi, M.; Khalid, M.S. UTM sustainable living laboratory campus; Are the implementations effective? In Proceedings of the Regional Conference in Engineering Education, Kuala Lumpur, Malaysia, 9–10 August 2016; pp. 1–6.
31. Humblet, E.M.; Owens, R.; Roy, L.P.; McIntyre, D.; Meehan, P.; Sharp, L. *Roadmap to a Green Campus*; U.S. Green Building Council: Washington, DC, USA, 2010.
32. Anthony, J.; Majid, M.A.; Romli, A. Emerging case oriented agents for sustaining educational institutions going green towards environmental responsibility. *J. Syst. Inf. Technol.* **2019**, *21*, 186–214. [CrossRef]
33. Baek, K.; Lee, H.; Chung, J.H.; Kim, J. Electric scooter sharing: How do people value it as a last-mile transportation mode? *Transp. Res. Part D Transp. Environ.* **2021**, *90*, 102642. [CrossRef]
34. Liu, M.; Seeder, S.; Li, H. Analysis of E-scooter trips and their temporal usage patterns. *Inst. Transp. Eng. ITE J.* **2019**, *89*, 44–49.
35. McKenzie, G. Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in Washington, D.C. *J. Transp. Geogr.* **2019**, *78*, 19–28. [CrossRef]
36. Kowald, M.; Gutjar, M.; Röth, K.; Schiller, C.; Dannewald, T. Mode Choice Effects on Bike Sharing Systems. *Appl. Sci.* **2022**, *12*, 4391. [CrossRef]
37. Eccarius, T.; Lu, C.-C. Adoption intentions for micro-mobility—Insights from electric scooter sharing in Taiwan. *Transp. Res. Part D Transp. Environ.* **2020**, *84*, 102327. [CrossRef]
38. Younes, H.; Zou, Z.; Wu, J.; Baiocchi, G. Comparing the temporal determinants of dockless scooter-share and station-based bike-share in Washington, DC. *Transp. Res. Part A Policy Pract.* **2020**, *134*, 308–320. [CrossRef]
39. Portland Bureau of Transportation. E-Scooter Findings Report. 2018. Available online: <https://www.portlandoregon.gov/transportation/article/709719> (accessed on 15 June 2018).
40. Denver Dockless Mobility Program. Pilot Interim Report—February 2019. Available online: <https://www.denverinc.org/wp-content/uploads/2019/05/Denver-Dockless-Mobility-Update-Feb-2019.pdf> (accessed on 1 February 2019).
41. The Nunatak Group. New Urban Mobility. 2019. Available online: <https://www.nunatak.com/en/topics/new-urban-mobility> (accessed on 20 July 2019).
42. 6t-Bureau de Recherche. Usages et Usagers des Trotinettes Electriques en Free-Floating en France. 2019. Available online: <https://6-t.co/etudes/usages-usagers-trotinettes-ff/> (accessed on 1 February 2019).
43. Sarker, I.H.; Colman, A.; Han, J.; Khan, A.I.; Abushark, Y.B.; Salah, K. BehavDT: A behavioral decision tree learning to build user-centric context-aware predictive model. *Mob. Netw. Appl.* **2019**, *25*, 1151–1161. [CrossRef]
44. Toraih, E.A.; Elshazli, R.M.; Hussein, M.H.; Elgaml, A.; Amin, M.; El-Mowafy, M.; El-Mesery, M.; Ellythy, A.; Duchesne, J.; Killackey, M.T.; et al. Association of cardiac biomarkers and comorbidities with increased mortality, severity, and cardiac injury in COVID-19 patients: A meta-regression and decision tree analysis. *J. Med. Virol.* **2020**, *92*, 2473–2488. [CrossRef]
45. Ganggayah, M.D.; Taib, N.A.; Har, Y.C.; Lio, P.; Dhillon, S.K. Predicting factors for survival of breast cancer patients using machine learning techniques. *BMC Med. Inform. Decis. Mak.* **2019**, *4*, 48. [CrossRef]
46. Lu, H.; Ma, X. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere* **2020**, *249*, 126169. [CrossRef] [PubMed]
47. Mosca, E.; Alfieri, R.; Merelli, I. A multilevel data integration resource for breast cancer study. *BMC Syst. Biol.* **2010**, *4*, 76. [CrossRef]
48. Genuer, R.; Poggi, J.-M.; Tuleau-Malot, C. VSURF: An R Package for Variable Selection Using Random Forests. *R J.* **2015**, *7*, 19–33. [CrossRef]
49. Lebedev, A.V.; Westman, E.; Van Westen, G.J.P.; Kramberger, M.G.; Lundervold, A.; Aarsland, D.; Soininen, H.; Kloszewska, I.; Mecocci, P.; Tsolaki, M.; et al. Random Forest ensembles for detection and prediction of Alzheimer’s disease with a good between-cohort robustness. *NeuroImage Clin.* **2014**, *6*, 115–125. [CrossRef]
50. Khalilia, M.; Chakraborty, S.; Popescu, M. Predicting disease risks from highly imbalanced data using random forest. *BMC Med. Inform. Decis. Mak.* **2011**, *11*, 51. [CrossRef]
51. Chen, S.; Webb, G.I.; Liu, L.; Ma, X. A novel selective naïve Bayes algorithm. *Knowl.-Based Syst.* **2020**, *192*, 105361. [CrossRef]

52. Ding, S.; Zhao, H.; Zhang, Y.; Xu, X.; Nie, R. Extreme learning machine: Algorithm, theory and applications. *Artif. Intell. Rev.* **2015**, *44*, 103–115. [[CrossRef](#)]
53. Tibshirani, R.; Walther, G.; Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2001**, *63*, 411–423. [[CrossRef](#)]
54. Nidheesh, N.; Nazeer, K.A.A.; Ameer, P.M. A Hierarchical Clustering algorithm based on Silhouette Index for cancer subtype discovery from genomic data. *Neural Comput. Appl.* **2020**, *32*, 11459–11476. [[CrossRef](#)]
55. Rai, P. Data clustering: K-means and hierarchical clustering. *CS5350 6350 Mach. Learn. Oct.* **2011**, *4*, 24.
56. Raschka, S. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. *arXiv* **2018**, arXiv:1811.12808.
57. Ramli, N.A.; Zen, I.S.; Bandi, M.; Tajuddin, H.A. Reduction in carbon dioxide emissions and global climate in campus: From policy into action. In Proceedings of the 2nd International Conference on Emerging Trends in Scientific Research, Kuala Lumpur, Malaysia, 1–2 November 2014; pp. 1–23.
58. Nejati, M.; Nejati, M. Assessment of sustainable university factors from the perspective of university students. *J. Clean. Prod.* **2013**, *48*, 101–107. [[CrossRef](#)]
59. Taghavi, M.; Bakhtiyari, K.; Taghavi, H.; Attar, V.O.; Hussain, A. Planning for sustainable development in the emerging information societies. *J. Sci. Technol. Policy Manag.* **2014**, *5*, 178–211. [[CrossRef](#)]
60. Foo, K.Y. A vision on the role of environmental higher education contributing to the sustainable development in Malaysia. *J. Clean. Prod.* **2013**, *61*, 6–12. [[CrossRef](#)]
61. Junior, B.A.; Majid, M.A.; Romli, A. Green information technology for sustainability elicitation in government-based organisations: An exploratory case study. *Int. J. Sustain. Soc.* **2018**, *10*, 20–41. [[CrossRef](#)]
62. Abdul-Azeez, I.A.; Ho, C.S. Realizing low carbon emission in the university campus towards energy sustainability. *Open J. Energy Effic.* **2015**, *4*, 15. [[CrossRef](#)]
63. Azlin, A.Z.B.; Er, A.C.; Rahman, N.B.A.; Alam, A.S.A. Consumers' roles and practices towards sustainable UKM campus. *Int. J. Adv. Appl. Sci.* **2016**, *3*, 30–34.
64. Peter, C.J.; Libunao, W.H.; Latif, A.A. Extent of education for sustainable development (ESD) integration in Malaysian community colleges. *J. Tech. Educ. Train.* **2016**, *8*, 1–13.
65. Junior, B.A. A retrospective study on green ICT deployment for ecological protection pedagogy: Insights from field survey. *World Rev. Sci. Technol. Sustain. Dev.* **2019**, *15*, 17–45. [[CrossRef](#)]
66. Hardt, C.; Bogenberger, K. Usage of e-Scooters in Urban Environments. *Transp. Res. Procedia* **2019**, *37*, 155–162. [[CrossRef](#)]
67. Laa, B.; Leth, U. Survey of E-scooter users in Vienna: Who they are and how they ride. *J. Transp. Geogr.* **2020**, *89*, 102874. [[CrossRef](#)]
68. Gössling, S. Integrating e-scooters in urban transportation: Problems, policies, and the prospect of system change. *Transp. Res. Part D Transp. Environ.* **2020**, *79*, 102230. [[CrossRef](#)]
69. Willmott, A.G.B.; Maxwell, N.S. The metabolic and physiological responses to scootering exercise in a field-setting. *J. Transp. Health* **2019**, *13*, 26–32. [[CrossRef](#)]
70. De Bortoli, A.; Christoforou, Z. Consequential LCA for territorial and multimodal transportation policies: Method and application to the free-floating e-scooter disruption in Paris. *J. Clean. Prod.* **2020**, *273*, 122898. [[CrossRef](#)]