SUSTAINABLE MOBILITY, TRANSIT-ORIENTED DEVELOPMENT, AND TRAVEL BEHAVIOR: A TRAVEL PATTERN ANALYSIS

AYA HASAN ALKHEREIBI
Department of Mechanical and Industrial Engineering, Qatar University, Qatar

ABSTRACT

This research proposes an inclusive travel pattern classification model, in support of sustainable mobility and transit oriented development to develop homogeneous activity groups for a sample of laborers in the State of Qatar. The investigated model aims to classify the activity data into a homogenous travel pattern. A pattern recognition model is applied to a revealed preference (RP) survey for the travel diary of 1,051 laborers. In the first phase of the analysis, raw data preprocessing algorithms and outliers data detection and filtering algorithms were applied and, therefore, an activity-based displacement matrix was developed for each household. The research methodology commenced in this research encompasses the integration of several machine learning (ML) techniques, mainly utilizing clustering and classification methods. A bagged clustering algorithm was used to recognize the clusters' number, and then the implemented CMeans algorithm and Pamk algorithm were used to validate the results. Meanwhile, the interdependencies between the resulting clusters and the socio-demographic characteristics of the household were examined using cross-analysis. The results of the study found that there was a notable diversity between clusters in terms of trip purpose, modal split, choice of destination, and occupation. Clustering techniques on all three attributes produced similar results, but clustering based on occupation yielded clusters that differed significantly from those based on other attributes. Applying such pattern recognition models to large and complex activity datasets could help transportation planners better understand the travel needs of segments of the population and formulate more informed strategies that compromise the best practices of sustainable mobility and transit-oriented development.

Keywords: travel behavior, transportation planning, activity-based model, machine learning, support vector machine, blue collar travel diary.

1 INTRODUCTION

1.1 Overview

In contemporary city planning, the use of public transportation is often considered one of the most significant strategies to reduce traffic congestion and pollution [1]. However, the snag of this approach is that people are less likely to walk around their neighborhoods and socialize. To counter this, transit-oriented development (TOD) has become one of the most popular urban planning approaches in recent decades [2]. This type of urban development aims to promote the use of mass transit by providing an environment where residents live within walking distance of a major through station and other amenities. It's expected that this will increase ridership and the use of public transport while allowing access to more jobs, education, and cultural facilities [3].

TOD is also aimed at providing a pedestrian environment and better accessibility, which promotes pedestrian activities [2]. Mentioning residents' activities, implement machine learning classifiers to recognize activity-based travel behavior patterns for a sample of laborers who represents a significant portion of residents of the Qatari society. This research problem is tackled from a dissimilar perception by developing a classification activity-based

WIT Transactions on The Built Environment, Vol 212, © 2022 WIT Press
www.witpress.com, ISSN 1743-3509 (on-line)
doi:10.2495/UMT220041

model that groups the homogeneous daily activity data into clusters. Each cluster has its essential features such as the type of activity, sequences, start time, activity duration, etc.

Moreover, the conventional trip-based travel demand modeling approach has been criticized for its conceptual deficiencies and lack of sensitivity to several important factors, including congestion, the distribution of activities over space, and the effects of economic, gender, social, transportation network, and locational divisions on spatial and temporal aspects of individual mobility [4]. The activity-based approach views travel as a derived demand; derived from the need to pursue activities distributed over space [5]. The approach focuses on sequences or patterns of activity behavior, with the whole day or longer periods of time as the unit of analysis. It emphasizes the effects of (and constraints imposed by) economic, gender, social, transportation network, and locational divisions on spatial and temporal aspects of individual movement.

The activity-based holistic perspective of travel can reliably evaluate urban travel demand management policies since it explicitly models activity patterns and considers these patterns to be the fundamental influence on travel. The objective of this ongoing research is to contribute toward advancing and operationalizing the activity-based approach by (a) Examining and understanding activity-based travel behavior, (b) developing a comprehensive representation of attributes characterizing laborers' daily activity-travel pattern, and (c) proposing an analysis framework for modeling the activity-travel pattern of individuals based on careful descriptive analysis of data, d) based on the defined travel pattern the research proposes TOD strategies that enhance smart mobility practices in the state of Qatar.

1.2 Literature review

Machine learning methods have been a boon to the transportation industry. They've aided planners in identifying patterns in travel data and using them to predict future travel behavior. Recently, Allahviranloo & Aissaoui [6] applied k-means clustering to identify distinctive travel clusters. They applied the AdaBoost algorithm to predict the start time and duration of travel-based activity based on individuals' sociodemographic features and individuals' daily schedules [7]. Researchers have used interpersonal variability to generate intrapersonal variability, but the new technique introduces a data-driven activity scheduling approach, which swaps several expert-designed mechanisms and their complex engineering interactions for a handful of machine learning techniques. Drchal et al. [8] noted that the new technique is more efficient than previous methods, as it does not require extensive engineering efforts or extensive data collection. They also noted that it is more accurate than previous methods, as it has a higher ability to predict human behavior. In order to execute this technique, the authors used an optimization algorithm based on a genetic algorithm (GA) and a particle swarm optimization (PSO). The resulting model is capable of generating an accurate schedule in real-time while being able to adapt dynamically to changes in traffic conditions or other external circumstances.

Activity-based travel demand models have made significant advances in the past decade, but there are still some significant opportunities for improvement. Despite advances in generating temporal information for the design of activity-based transport demand models, there is significant scope and opportunity to improve the performance of these models in terms of estimation accuracy, estimation efficiency, computations, and their impact on case studies.

The activity-based modeling is a technique that can be used to identify patterns in the activities of people on a daily basis. This model has been utilized by several scholars such as

Hafezi et al. [9] and Awais Shafique and Hato [10]. The key idea behind activity-based modeling is that it can be used to model the temporal dimensions allied with the traveler's daily activity patterns, which have an important impact on the travel experience. The random forest algorithm was proposed by Hafezi et al. [9]. It uses a random forest model with CART techniques for measuring features' importance. However, this technique has also been used in other areas such as traffic accident detection and means of transport [10]. In their study, using activity and temporal models, Hafezi et al. [9] improved current activity-based techniques by using daily data and pattern recognition models to identify groups of people who have ongoing demands for their activities outside of their place of residence. And more broadly, similar travel patterns along with the type of activities and start/end times along with the duration.

The SALT planner framework uses a set of advanced rule-based econometric models to predict daily trip demand specifications and route generation. Identifying homogeneous populations can significantly improve the estimation accuracy of activity-based transport demand models. Moreover, the SALT planner framework is designed to obtain a dynamic model for traffic demand in an urban area by using both time-series data and spatial information. It uses an advanced combination of traffic forecasting techniques through an econometric model to predict daily trip demand specifications and route generation.

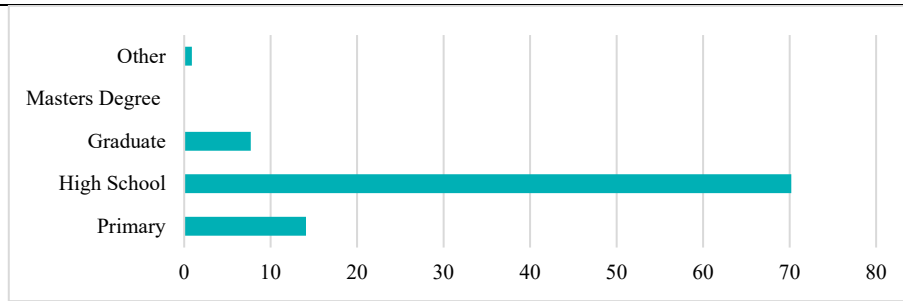
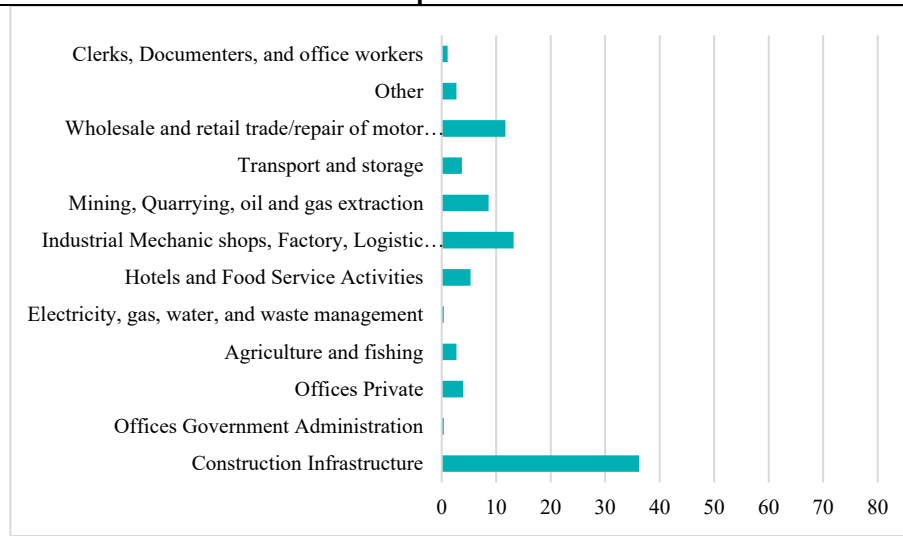
2 METHODS AND MATERIALS

Table 1 shows the general characteristics, in terms of age, income, education level, and occupation, of the investigated sample.

Table 1: Sample general characteristics.

Sample characteristics	
Age	
55 _ 74 Years old	
45 _ 54 Years old	
35 _ 44 Years old	
25 _ 34 Years old	
0 5 10 15 20 25 30 35 40 45 50	
Income	
10K _ 20K QR/M	
6K _ 10K QR/M	
3K _ 6K QR/M	
1K _ 3K QR/M	
1K QR/M	
0 10 20 30 40 50 60 70	

Table 1: Continued.

Sample characteristics
Education

Occupation sector

Occupation level


Age: More than 93% of the respondents were from the category of economically active ages ranging from 19 to 54. Similarly, the overwhelming majority of the respondents (93%) earn QAR3,000 ($USD 820) or less per month. The education level of respondents is mostly secondary and high school educators with a total percentage of 70.2.

Income: Almost all (61%) respondents indicated that they earn no more than QAR3,000 per month while 3% earn between QAR3,001–6,000 per month; 1% earn between QAR6,001–10,000 per month, and 0.3% earn more than QAR10,001 per month.

Activity-based travel patterns (ABTP) are travel recognition pattern recognition that is based on activities that may change over time. The recognition and comprehension of activity-based travel patterns are important for developing efficient transportation systems [6], [11]–[13]. One method used by Leisch [14] was the bagged clustering (B-Cluster) algorithm for initialization of the number of clusters. Subsequently, C-Means and Pamk algorithms were applied to optimize the performance of the B-Cluster algorithm as recommended by Hung and Yang [15]. In this study, we used the CROSSTABS test to deal with interdependencies and applied the CART classifier algorithm for the alignment of a person's socio-demographic characteristics with predefined clusters. The theoretical model proposed herein has the capability of learning and predicting temporal information of patterns of activity for individuals concerning the characteristics of heterogeneity. Besides, applying this framework to activity-based models reveals the effectiveness of machine learning techniques in distinguishing temporal attributes of travelers' daily activity patterns, but also helps better understand the relationship between travel planning and patterns of activities with respect to heterogeneity characteristics.

A flowchart of the research methodology followed in this study is displayed in Fig. 1.

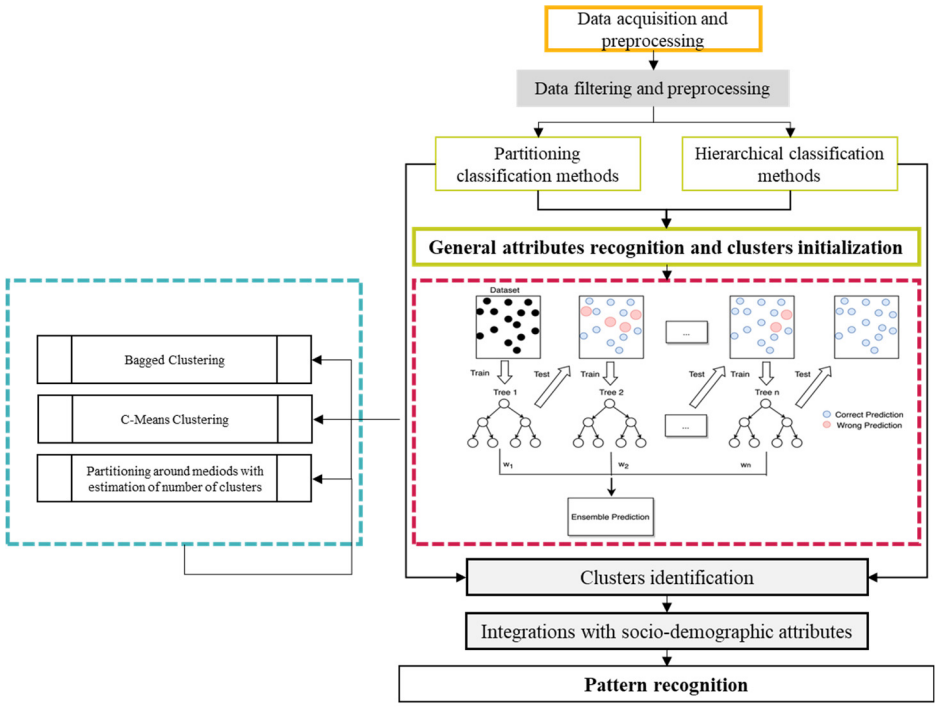


Figure 1: Research design flowchart.

3 RESULTS AND KEY FINDINGS

Different structures are observed by applying different grouping techniques. Table 2 describes the results of these clustering techniques. Figs 2–4 show the plots for the three clustering techniques. As described, it appears that Cmeans and Pamk produce an even distribution compared to the bag clustering technique. The latter yields insignificant records for clusters 5, 6, 7, 9, and 10, as well as null Logs for Cluster 8.

Table 2: Number of clusters per method.

Method	Cluster										
	1	2	3	4	5	6	7	8	9	10	Total
Bclust	2,150	2,405	28	356	32	3	4	0	8	4	4,990
Cmeans	358	583	43	658	353	612	404	1,369	572	38	4,990
Pumk	604	1692	693	443	591	529	387	38	8	5	4,990

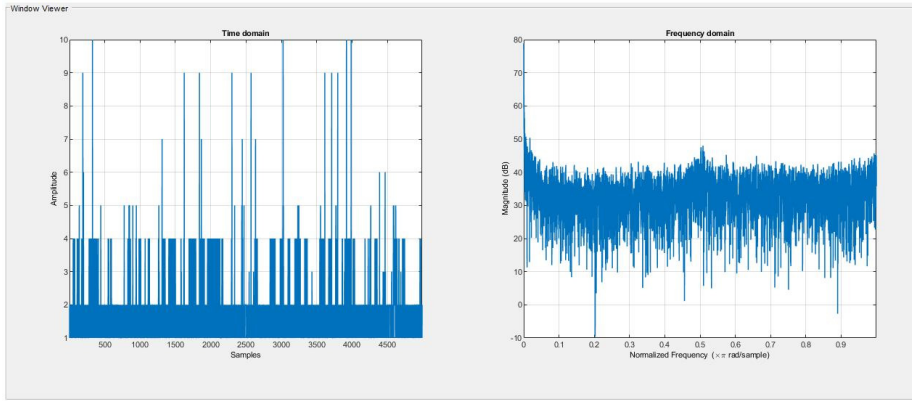


Figure 2: Bagged clusters.

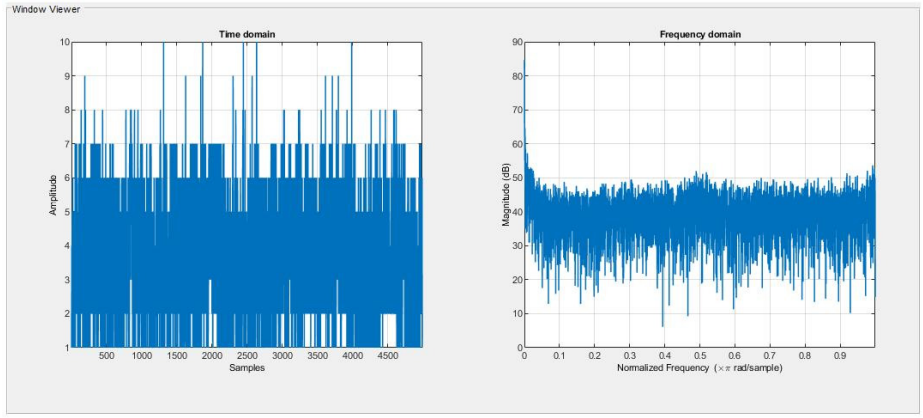


Figure 3: C-means clusters.

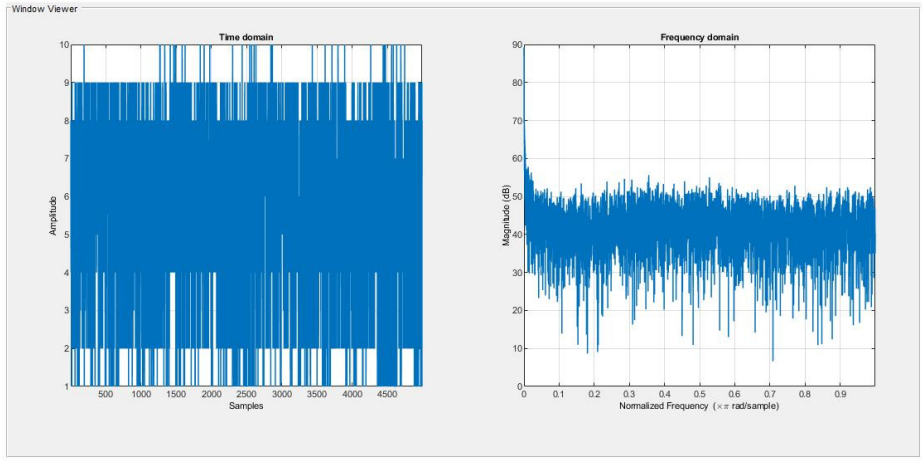


Figure 4: Pamk clusters.

The C-means clustering technique has been used to cluster the activities of laborers based on their work, job duties, and personal activities. According to the results, it is observed that the three main activities are work and job duties, personal activities, and preparing meals for a group of co-workers. These have shares of 21%, 43%, and 31% respectively. Engagement in other activities such as study is negligible. Then the Bagged clustering technique was used to create a set of clusters from observations with similar attributes; this is done by combining several methods of clustering into one. This combines the advantages of each individual method and reduces the disadvantages of each individual method. Utilizing the bagged clustering gave the same characteristics of cluster 1 as the C-means technique.

In the context of traveling to work, it is possible to make a distinction between three main modes: returning home, commuting to the workplace and traveling to other places. By the way, Al Rayyan, Al Sheehaniya and Doha are respectively three main municipalities where trips were made.

Al Rayyan is the municipality with the highest share of trips in all categories (i.e. working trips, leisure and shopping trips). On the other hand, Al Sheehaniya is second in terms of transportation modes (i.e. private shuttle and car/van/truck) whereas Doha is ranked third for both categories (i.e. private shuttle and car/van/truck).

In terms of income, the majority of earners make between QR 1,001 and QR 3,000 per month (around 63%). Around 33% earn less than QR 1,000 monthly; 2.5% earn between QR 3,001 and QR 6,000 per month; only 1% earns more than QR 6,000 monthly every month.

The majority of the workers in this cluster were Indian (43%) followed by Nepalese, Bangladeshi and Filipino nationals respectively. Around 40% of the workers were between 25 and 34 years old. The other demographics include a high percentage of male workers at 98.9% and full-time employed adults at 99.9%. Nearly 84% of all occupations are classified as elementary occupations such as drivers and laborers while 9% are factory or industrial workers such as plant/machine operators and assemblers. Approximately 3% are technicians and associate professionals.

The main domain for these occupations is construction infrastructure (30%), industrial mechanics shops (20%), logistics manufacturing (10%), mining quarrying oil gas extraction (10%) motor mechanics (10%). Other domains include agriculture fishing forestry social

services utilities transportation communications commerce finance insurance real estate management consulting professional scientific technical arts entertainment recreation hospitality.

According to the results presented above, the dominant transport mode used by laborers is the private shuttle. As stated by Shaaban and Maher [16], Qatar is one of the highest countries in private car ownership, these facts created the emergence to finding more sustainable solutions such as sustainable mobility and transit oriented development. Moreover, this study shows that there is an urgent need for studying the feasibility of implementing sustainable mobility and transit-oriented development in order to increase accessibility to public transportation, especially in the presence of the poor use of PT; less than (3%). Additionally, with the huge urban and infrastructure evolution Qatar witnessed in the preparation for FIFA World Cup 2022, the country had a huge development in public transportation as well. In this context, several sustainable mobility strategies could be employed as suggested by Hafezi et al. [13] to help mitigate the unsustainable travel pattern. These strategies are represented under different categories:

- Transportation: by increasing the use of public transportation, carpooling, and ride sharing, and reducing the distance between home and work by the well-planned allocation of laborers camps.
- Infrastructure: includes developing more walkable communities with bike trails and other amenities that encourage walking and biking. It also includes converting unused spaces into bicycle parking areas and creating better bicycle paths, and providing a shaded path to enhance walkability.
- Cultural awareness: by educating the public and especially laborers about the importance of sustainable mobility practices through campaigns.

4 CONCLUSION

This research presented the applicability of machine learning classifier models to an activity-based model for a sample of laborers in Qatar, to understand laborers travel patter then help decision-makers formulate strategies for the best practices of sustainable mobility and transit-oriented development. The proposed model applied clustering techniques to consolidate household activity data into a homogeneous activity model.

This study was undertaken to extract socio-economic and household travel information from a revealed preference (RP) survey collected by the Ministry of Transport and Communications (MOTC) of Qatar, which was used to uncover patterns of residents' activity, particularly in the context of trip purpose, mode choice and destination. Three clustering techniques (bagged clustering, C-means clustering, and Pamk clustering) were applied separately to uncover patterns of residents' activity in terms of trip purpose, mode choice and destination. The findings from this research study prove that the above three clustering methods are credible and could provide a better understanding of the residents' shift pattern. However, the CMeans and Pamk clustering technique show better efficiency and representativeness than the bagged clustering technique. The study clustered a large number of activities collected from travel diaries into meaningful clusters using machine learning techniques to best support the further development of predictive models for diverse elements of the transportation system Travel patterns have been identified which help to develop effective decision support systems in transit-oriented development planning and evaluation of different strategies. The results of the study showed that cluster analysis techniques are mathematically efficient and can classify residents into groups and, therefore, analyze travel behavior. By applying clustering techniques, detailed traveler characteristics

such as income, gender, age, nationality and occupation are considered for activity-based models in addition to mode/destination choice. The identified travel patterns are valuable know-how and absorption in the development of effective decision-making systems in the sustainable mobility and transit-oriented development planning and evaluation of different strategies.

REFERENCES

[1] Chan, W.C., Ibrahim, W.H.W., Lo, M.C., Suaidi, M.K. & Ha, S.T., Sustainability of public transportation: An examination of user behavior to real-time GPS tracking application. *Sustainability (Switzerland)*, 12(22), pp. 1–20, 2020. DOI: 10.3390/su12229541.

[2] Thomas, R., Pojani, D., Lenferink, S., Bertolini, L., Stead, D. & van der Krabben, E., Is transit-oriented development (TOD) an internationally transferable policy concept? *Reg. Stud.*, **52**(9), pp. 1201–1213, 2018. DOI: 10.1080/00343404.2018.1428740.

[3] Nasri, A. & Zhang, L., The analysis of transit-oriented development (TOD) in Washington, D.C. and Baltimore metropolitan areas. *Transp. Policy (Oxf.)*, **32**, pp. 172–179, 2014. DOI: 10.1016/j.tranpol.2013.12.009.

[4] Moeckel, R., Kuehnel, N., Llorca, C., Moreno, A.T. & Rayaprolu, H., Agent-based simulation to improve policy sensitivity of trip-based models. *J. Adv. Transp.*, **2020**, 2020. DOI: 10.1155/2020/1902162.

[5] Joubert, J.W. & de Waal, A., Activity-based travel demand generation using Bayesian networks. *Transp. Res. Part C Emerg. Technol.*, **120**, 102804, 2020. DOI: 10.1016/j.trc.2020.102804.

[6] Allahviranloo, M. & Aissaoui, L., A comparison of time-use behavior in metropolitan areas using pattern recognition techniques. *Transp. Res. Part A Policy Pract.*, **129**, pp. 271–287, 2019. DOI: 10.1016/j.tra.2019.08.007.

[7] Li, Y., Wang, C. & Gong, J., A multi-model probability SOC fusion estimation approach using an improved adaptive unscented Kalman filter technique. *Energy*, **141**, pp. 1402–1415, 2017. DOI: 10.1016/j.energy.2017.11.079.

[8] Drchal, J., Čertický, M. & Jakob, M., Data-driven activity scheduler for agent-based mobility models. *Transp. Res. Part C Emerg. Technol.*, **98**, pp. 370–390, 2019. DOI: 10.1016/j.trc.2018.12.002.

[9] Hafezi, M.H., Daisy, N.S., Millward, H. & Liu, L., Ensemble learning activity scheduler for activity based travel demand models. *Transp. Res. Part C Emerg. Technol.*, **123**, 102972, 2021. DOI: 10.1016/j.trc.2021.102972.

[10] Awais Shafique, M. & Hato, E., Use of acceleration data for transportation mode prediction. *Transportation*, **42**, pp. 163–188, 2015. DOI: 10.1007/s11116-014-9541-6.

[11] Cui, Y., He, Q. & A. Khani, A., Travel behavior classification: An approach with social network and deep learning. *Transp. Res. Rec.*, **2672**(47), pp. 68–80, 2018. DOI: 10.1177/0361198118772723.

[12] Chen, S., Prakash, A.A., de Azevedo, C.L. & Ben-Akiva, M., Formulation and solution approach for calibrating activity-based travel demand model-system via microsimulation. *Transp. Res. Part C Emerg. Technol.*, **119**, 102650, 2020. DOI: 10.1016/j.trc.2020.102650.

[13] Hafezi, M.H., Daisy, N.S. & Millward, H. & Liu, L., Ensemble learning activity scheduler for activity based travel demand models. *Transp. Res. Part C Emerg. Technol.*, **123**, 102972, 2021. DOI: 10.1016/j.trc.2021.102972.

[14] Leisch, F., Bagged clustering. *Adaptive Information Systems and Modelling in Economics and Management Science*, **51**(51), p. 11, 1999. DOI: 10.1.1.33.642.

[15] Hung, M. & Yang, D., *An Efficient Fuzzy C-Means Clustering Algorithm Ming-Chuan,* IEEE, 2001.
[16] Shaaban, K. & Maher, A., Using the theory of planned behavior to predict the use of an upcoming public transportation service in Qatar. *Case Stud. Transp. Policy*, **8**(2), pp. 484–491, 2020. DOI: 10.1016/j.cstp.2019.11.001.