



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Transport & Health

journal homepage: www.elsevier.com/locate/jth

The identification of significant features towards travel mode choice and its prediction via optimised random forest classifier: An evaluation for active commuting behavior

Nur Fahriza Mohd Ali^{a,*}, Ahmad Farhan Mohd Sadullah^a,
Anwar P.P. Abdul Majeed^{b,d}, Mohd Azraai Mohd Razman^b, Rabiuh Muazu Musa^c

^a School of Civil Engineering, Engineering Campus, Universiti Sains Malaysia, 14300, Nibong Tebal, Pulau Pinang, Malaysia

^b Innovative Manufacturing, Mechatronics and Sports Laboratory, Faculty of Manufacturing and Mechatronics Engineering Technology, Universiti Malaysia Pahang, 26600, Pekan, Pahang, Malaysia

^c Center for Fundamental and Continuing Education, Universiti Malaysia Terengganu, Kuala Nerus, Terengganu, Malaysia

^d School of Robotics, XJTU Entrepreneur College (Taicang) Xi'an Jiaotong-Liverpool University, Suzhou. 215123. P. R. China

ARTICLE INFO

Keywords:

Door-to-door journey
Travel mode choice
Random forest classifier
Public transport
Private vehicles
Hyperparameter tuning

ABSTRACT

Introduction: Physical activity is the foundation to staying healthy, but sedentary activities have become not uncommon that ought to be mitigated immediately. The study aims to highlight the role of a transport system that encourages physical activity among users by applying an active door-to-door transport system. Users' mode choice is studied to understand their preferences for active commuting. The use of machine learning has since been ubiquitous in a myriad of fields, including transportation studies and hence is also investigated towards its efficacy in predicting travel mode choice.

Methodology: The application of the Random Forest (RF) model to identify travel mode choice is explored using the Revealed/Stated Preferences (RP/SP) Survey data in Kuantan City during weekdays. A total of 386 respondents were involved in this survey. The efficacy of the tuned RF models towards predicting the travel mode choice is evaluated via the Classification Accuracy (CA) performance indicator. In addition, a Feature Importance study is also carried out in order to identify significant factors that contribute towards travel mode choice.

Results: The results from the present investigation demonstrated that the default RF model has acceptable predictability for both training and test dataset of users' mode choice, with a CA of 70.2% and 69.3%, respectively. Upon identifying the significant features and further refining the hyperparameters of the RF model heuristically, it was shown that with 145 trees, the CA improved to up to 71.6% and 70.1% for both the training and test dataset, respectively. Through the feature selection technique, the most significant features that affect users mode choice are total travel time (TT), waiting time at a public transport stop (WT), region, walking distance from the last stop to destination (WD2), and walking distance from home to the nearest bus stop (WD1).

Conclusions: The study has illustrated the efficacy of the optimised RF in predicting travel mode choice as well as identified the significant factors for the selection. The findings of the present study provide significant insight for policymakers to improve the performance of the public

* Corresponding author.

E-mail address: fahriza90@gmail.com (N.F. Mohd Ali).

<https://doi.org/10.1016/j.jth.2022.101362>

Received 2 April 2021; Received in revised form 13 January 2022; Accepted 10 March 2022

Available online 1 April 2022

2214-1405/© 2022 Elsevier Ltd. All rights reserved.