Predicting mode choice using boosted trees in a multi-level panel effect model

Pim Labee*[1], Soora Rasouli[1], and Seheon Kim[2]

[1]Design and Decision Support Systems: Urban Planning and Transportation group, Department of the Built Environment, Eindhoven University of Technology
[2]Department of Urban Engineering, Hanbat National University, South Korea

SHORT SUMMARY

Predicting travel mode choice is a crucial aspect of transportation planning and research. It involves developing models and methodologies to anticipate the mode of transportation individuals are likely to choose for a given trip. Both discrete choice models and machine learning techniques are often used to analyze historical travel behavior data and derive patterns that can be used for prediction. These models help urban planners and policymakers make informed decisions about transportation infrastructure, public transit services, and sustainable mobility options. Both discrete choice models and machine learning models have strengths and weaknesses. In this paper, we present a method that is able to harness the strengths of advanced gradient boosted decision trees while accounting for the panel structure in the data and estimating random effects, which - in machine learning studies - are otherwise often ignored. The models are tested on a travel mode choice case study and show improved predictive performance compared to a plain gradient boosted decision trees model.
Keywords: activity-based modelling, boosted tree, panel effects

1 INTRODUCTION

Travel mode choice often guides travel policies and decisions in urban planning. Therefore it's not surprising that travel mode choice is one of the core elements of travel demand modelling. Discrete choice models have long dominated the literature being the most suitable technique to model mode choice. Due to the increasing availability of data and computational power, machine learning models have been increasingly applied to predict mode choice. The data-driven nature of machine learning models allow for more flexibility than discrete choice models, resulting in better predictive performance (Zhao et al., 2020). The decision tree (Breiman et al., 1984), and later the random forest (Breiman, 2001) and boosted trees algorithms (Ke et al., 2017), are among the most popular machine learning algorithms for predicting mode choice both because of their interpretability as well as predictive performance (Hagenauer and Helbich, 2017). Even though predictive performance of machine learning (ML) models are unparalleled by discrete choice models, ML models still generally fall short when it comes to modelling panel effects in repeated observations (Kim et al., 2018). Ignoring the presence of panel and/or autoregressive effects may lead to bias in the estimated effects, and thus predictive outcomes (Sela and Simonoff, 2012). Several efforts have been made to incorporate random effects into regression trees (Hajjem et al., 2011; Sela and Simonoff, 2012) and later regression random forests (Hajjem et al., 2014). The proposed algorithms for the random effects regression trees followed comparative structures, which resemble an expectation-maximization algorithm. In the first step, a normal regression tree is fit to the dependent variable minus the random effects. Initially the random effects are assumed to be zero. The nodes of the regression tree are then used as fixed effects in the estimation of a mixed effects model in the second step of the algorithm. These steps are iteratively conducted until convergence. Variations of this model for binary classification and count outcomes have then been proposed (Hajjem et al., 2017; Fokkema et al., 2018, 2021). Only few studies, however, have proposed a method to predict unordered categorical dependent outcomes (Kim et al., 2018), and to our knowledge, there are no studies that propose an algorithm to estimate random effects while using random forests or boosted trees. As random forests and boosted trees typically exhibit superior predictive performance compared to decision trees, incorporating random effects in them is a worthwhile endeavor. In light of the above arguments, this work proposes a method to predict the choice of transport mode while accounting for panel effects. In the second section the proposed methodology is highlighted. The results are subsequently discussed in section three, and conclusions are provided in the final section.

2 Methodology

The goal is to predict the main travel mode used by travelers for each trip, considering potential unobserved systematic behavior manifested in multiple choices of a single respondent, employing gradient boosting decision trees. In line with common practice, the main mode is defined as the mode that is used for the trip leg covering the longest distance. If multiple responses exist for the same respondent in longitudinal data and if not accounted for in the modelling, the assumption of independent measures would be violated. To account for the pattern that may exist in the behavior of the same individual, the mixed model has been developed (Equation 1) (Laird and Ware, 1982). Assuming we have a panel of individuals i=1,... I, making choices at times t=1,... T_i and were the functional form $f(x)$ assumed to be linear, one would arrive at the commonly used linear mixed model for continuous outcomes, as shown in Equation 2.

$$\boldsymbol{y}_i = f(x_{it1}, \ldots, x_{itp}) + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{1}$$

$$\begin{gathered} \boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \\ \begin{bmatrix} \boldsymbol{b}_i \\ \boldsymbol{\epsilon}_i \end{bmatrix} \sim N\left(0, \begin{bmatrix} D & 0 \\ 0 & R_i \end{bmatrix}\right) \end{gathered} \tag{2}$$

where $\boldsymbol{y}_i$ is the T_i x 1 column vector of continuous responses for the T observations of respondent i, $\boldsymbol{X}_i$ is the T_i x p matrix of the p fixed effect covariates, $\boldsymbol{\beta}$ is a p x 1 vector of the parameters to be estimated, $\boldsymbol{Z}_i$ is the T_i x q matrix of random effects covariates, $\boldsymbol{b}_i$ is the q x 1 vector of random effects, ε_i is the T_i x 1 column vector of errors and $\boldsymbol{D}$ and $\boldsymbol{R}_i$ are the covariance matrix of the random effects and errors respectively. Assuming a linear form, however, may be too restrictive since the functional form of the fixed effects is usually unknown (Sela and Simonoff, 2012). To alleviate this restrictive assumption, various extensions have been proposed, as discussed in the introduction. One approach involves estimating the fixed effects through non-parametric decision trees (Fokkema et al., 2018, 2021; Hajjem et al., 2011; Sela and Simonoff, 2012). When the observations $\boldsymbol{y}_i$ and random effects $\boldsymbol{b}_i$ are known, $f(x_{i1}, ..., x_{ip})$ can be approximated by fitting a regression tree to $\boldsymbol{y}_i - \boldsymbol{Z}_i \hat{\boldsymbol{b}}_i$. Conversely, if the observations $\boldsymbol{y}_i$ and the fixed effects $f(x_{i1}, ..., x_{ip})$ are known, $\boldsymbol{b}_i$ is estimated by fitting a traditional linear mixed model. However, when dealing with an unordered multinomial categorical outcome, such as mode choice, subtracting $\boldsymbol{Z}_i \hat{\boldsymbol{b}}_i$ from $\boldsymbol{y}_i$ is not feasible. A specific method (Kim et al., 2018) was proposed to deal with unordered multinomial categorical outcome, and further tested for mode choice in (Labee et al., In press.). In this work we propose an extension of the method to enable leveraging the strengths of boosted trees, while accounting for repeated choices nested across individuals. A gradient boosting decision trees model (GBDT) is initially fitted to multinomial outcome y_{it} taking values 1,... J, assuming the random effects are known. In previously proposed methods where a decision tree was used in the initial step, the leaves of the trees would be used as the fixed effects in the next step. Due to the large number of trees in random forests or boosting algorithms, this would be highly impractical. Instead, after the boosted trees model is fitted to the training data, the shapley values are extracted (Lundberg and Lee, 2017). These shapley values are then used as the input features for a k-means clustering model (MacQueen, 1967). The resulting clusters then form the fixed effects $\hat{f}_j(\boldsymbol{X}_i)$. The determined fixed effects can be employed in the subsequent phase to calculate the random effects, wherein different levels can be identified, as shown in Figure 1. A Bayesian regression model is applied to Equation 3, incorporating non-informative prior distributions. The fixed effects are assigned a prior normal distribution, while half Cauchy priors are employed for the random effects. Given the multinomial nature of the dependent variable, the baseline-category logit model (Agresti, 2002) is chosen as the suitable link function for the Generalized Linear Mixed Model (GLMM). Then the logits η_{ij} of each of the $J-1$ categories over the reference category, can be written as Equation 3.

$$\boldsymbol{\eta}_{ij} = \log \frac{\boldsymbol{\pi}_{\mathbf{ij}}}{\boldsymbol{\pi}_{\mathbf{iJ}}} = \hat{f}_j(\boldsymbol{X}_i) + \boldsymbol{Z}_i \boldsymbol{b}_{ij}, j = 1, ..., J-1 \tag{3}$$

where π_{ij} are the response probabilities for the non-baseline categories ($j \neq J$), and π_{iJ} the response probability for the baseline-category. The response probabilities without random effects can then be extracted as

$$\pi_{ijt}^{MCMC} = \frac{\exp(\hat{\eta}_{ijt} - z_{ijt} \hat{\boldsymbol{b}}_{ij})}{1 + \sum_{j'=1}^{J-1} \exp(\hat{\eta}_{ij't} - z_{ij't} \hat{\boldsymbol{b}}_{ij})}, j = 1, ..., J-1 \tag{4}$$

$$\pi_{ijt}^{MCMC} = \frac{1}{1 + \sum_{j'=1}^{J-1} \exp(\hat{\eta}_{ij't} - z_{ij't} \hat{\boldsymbol{b}}_{ij})}, j = J. \tag{5}$$

The response can be subsequently updated using the Monte Carlo technique from the probabilities π_{ijt}^{MCMC}. A thorough gridsearch was conducted to find the most suitable parameters to grow the gradient boosted trees. To have a decent effective sample size after the MCMC sampling, we have simulated four chains for 2,000 iterations with a warmup of 1,000 iterations. Thus, we ended up with 4*(2,000-1,000) post-warmup draws. All analyses have been conducted using R (R Core Team, 2022), using the LightGBM and BRMS packages (Shi et al., 2023; Bürkner, 2017).

Algorithm 1: Random effect gradient boosted tree estimation procedure

Input:
y- vector with responses y_{ij}
cov- data frame with all covariates
gr- vector of the grouping variable for each observation
$znam$- vector with names of covariates to be used as random effects
$xnam$- vector with names of covariates to be used as fixed effects
tol- threshold of convergence
$itmax$- maximum number of iterations

```
Set the estimated random effects b̂_ij = 0 and set the number of iterations n = 0
while n < itmax and not converged do
    Fit Gradient Boosting Decision Trees to y_it if n = 0, otherwise to y*_it
    Extract the shapley values
    Cluster observations using K-means algorithm
    Set n = n+1, and set clusters from previous step to f̂_j(X_i)
    Fit η_ij = f̂_j(X_i) + Z_i b_ij using Markov-Chain Monte-Carlo GLMM
    Extract b̂_ij from the estimated MCMC GLMM
    Update a response from the probabilities π̂^MCMC_ijt, and set it to y*_it
    if ΔWAIC < tol then
        converged ← true
    else
        converged ← false
    end
end
```

Output: The final fitted MCMC GLMM
The final fitted GBDT
$\boldsymbol{b}$, the final estimation of the random coefficients
n, the number of iterations
WAIC, the model fit of the mixed model

The proposed algorithm is shown in algorithm 1. Tested parameters and their optional values are shown in Table 1. Using the same set of parameters for the boosted trees model, we specify four different random intercept models. Two two-level models with a random intercept for each individual - one with 100 clusters and one with 200 clusters (model 1 and 3 respectively). These models are then compared to the plain boosted trees model and two three-level models, model 2 and 4 respectively, (also with 100 and 200 clusters respectively), where a random intercept for 'travel-day' within each individual is also considered. Travel-day represents an indicator for one of the three recorded travel-days by the individual, on which the trip was conducted.

The estimation utilizes data from the Netherlands Mobility Panel (MPN) travel survey (Hoogendoorn-Lanser et al., 2015). This dataset encompasses details at household, individual, and trip levels, with each respondent maintaining a travel diary for three consecutive days annually. A descriptive summary is shown in Table 2. Built environment characteristics were also included as part of the conditional variables. The incorporation of built environment features is motivated by prior research (Cervero, 2002; Cervero and Kockelman, 1997; Cheng et al., 2019; Ewing and Cervero, 2010; Kim et al., 2021), suggesting their impacts on the mode choice decision of travelers. This study includes factors such as distances to public transportation (PT) services such as train stations, bus, metro, and tram stops, and the city center. Land use mix of the respondent's home postal code zone is also considered, calculated as $\Sigma_m(P_m ln(P_m)/ln(k))$ where P_m is the proportion of the m-th land use category (Song et al., 2013) and k is the number of land

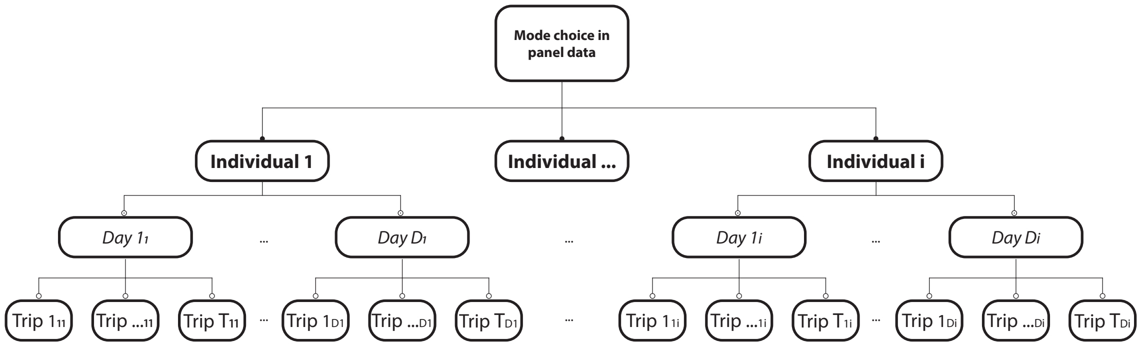

Figure 1: Hierarchy in the panel data

Table 1: Grid search parameters

Maximum depth	6, 8, 10, 20
Learning rate	0.01,0.05,0.1
Number of leaves	30, 40, 50
Minimum number of observations per leaf	10, 20, 50
Number of boosting iterations	1000
Use 'extraTrees' algorithm	Yes, no
Rounds for early stopping	50
Fraction of features used per tree	0.6, 0.8, 1.0
Fraction used for bagging	0.6, 0.8, 1.0
Bagging frequency	1, 5

use types *. Additionally, weather data, known to impact mode choice (Böcker et al., 2013, 2016; Hyland et al., 2018; Kim, 2020; Liu et al., 2015, 2017), are incorporated. The variables include daily mean wind speed, daily mean temperature, precipitation duration, daily precipitation amount, and the maximum hourly precipitation amount, collected from 29 weather stations in the Netherlands and geographically matched with the trip's origin. These variables only served as conditional variables, and no covariates were added. A stratified split is used to split the data into a training (80%) and testing data set (20%).

Table 2: Descriptive summary of the data

Time span	2018
Number of respondents	3,443
Number of (main) trips	33,443
Total number of trips per individual across all three days	Minimum = 1, first quartile = 6, median = 9, third quartile = 12, maximum = 42
Bicycle Bus Tram Metro (BTM) Car as driver (CaD) Car as passenger (CaP) E-bike Train Walking	20.8% 1.9% 39.4% 9.5% 6.4% 2.7% 19.3%

3 Results and discussion

In order to speed up the model estimation, the algorithm here is limited to one iteration. Figure 2 displays trace and density plots for the random effects (the random intercept at individual level) for model 4 (three-level model with 200 clusters), allowing us to assess the proper mixing of Markov chains during model estimation. The various plots indicate effective mixing of the chains: the aligned density plots and the distinctive 'hairy caterpillar' pattern in the trace plots on the right-hand side demonstrate this phenomenon (Lee and Wagenmakers, 2013). Table 3 shows the estimated random effects for the four models. Elevated standard deviation values across all models suggest a significant amount of variability in mode choice for

*included land use types: education, offices, stores, meeting (such as restaurants and bars), lodging and residential individuals' main trips. This underscores the importance of considering the nested nature of choices within individuals. The estimates for the random effects in model 1 are very similar to those in model 3, and therefor the increase in clusters in the k-means clustering step does not seem to affect those estimates. The same holds for models 2 and 4. In Table 4 we compare the predictive out of sample performance of the mixed effects models, conditionally on the estimated random effects and evaluate them against the predictive performance of the plain boosted trees model. Due to the class imbalance in the target variable, we used balanced accuracy (Equation 6), F1 score (Equation 7) and Cohen's Kappa (Equation 10), where the latter controls for the expected (random) accuracy.

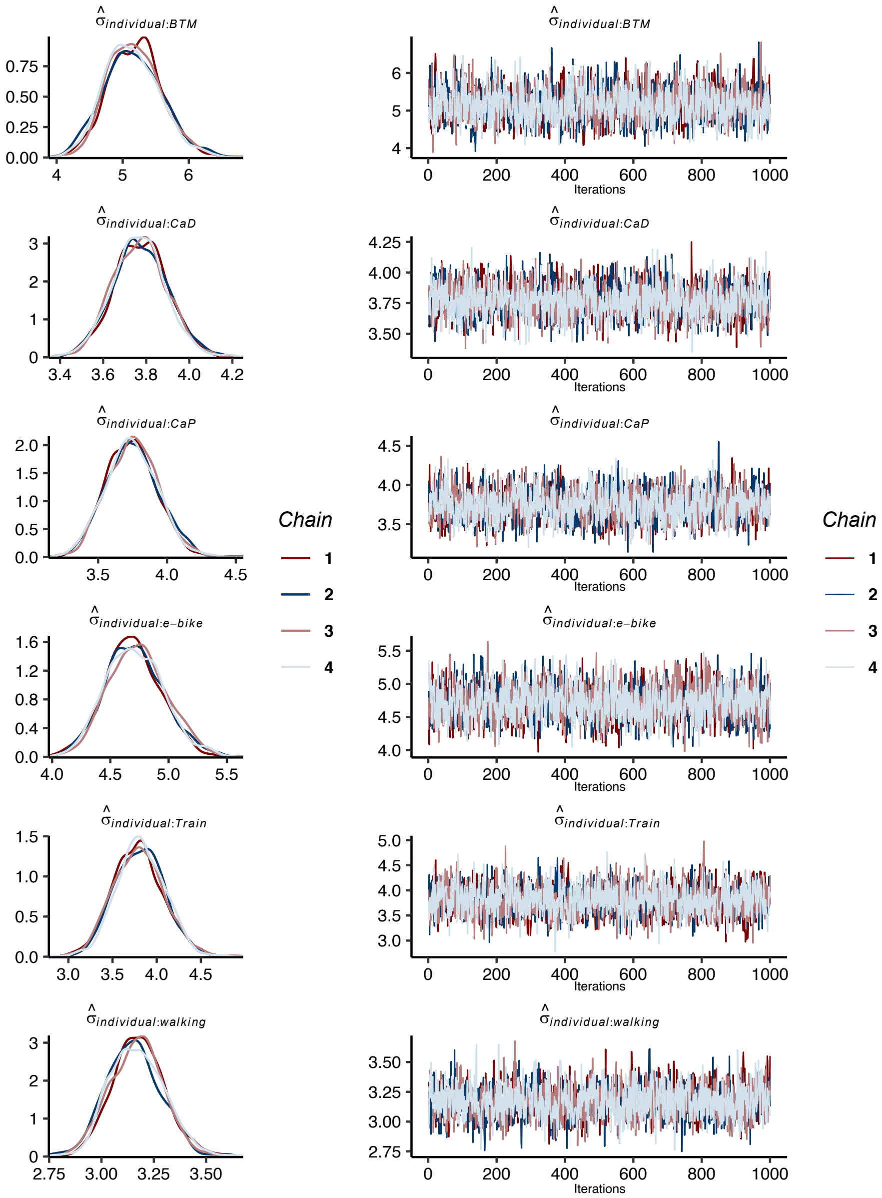

Figure 2: Trace and density plot random effects (individual level) (bicycle is reference category)-model 4

$$\text{balanced accuracy} = \frac{\frac{TP^{\dagger}}{TP+FN} + \frac{TN}{TN+FP}}{2} \tag{6}$$

$$F_1 = 2 * \frac{Precision * recall}{Precision + recall} \tag{7}$$

$$recall = \frac{TP}{TP + FN} \tag{8}$$

†TP=True Positive, TN = True Negative, FP = False Positive and FN = False Negative

Table 3: Random effects estimates (reference category is bicycle)

	BTM	CaD	CaP	e-bike	Train	Walking
Model 1	100 clusters					
Mean	0.745	0.448	0.557	0.681	0.778	0.447
$\sigma_{individual}$	4.263	2.903	3.541	3.930	4.169	2.717
Sd c.i.	3.759-4.813	2.748-3.060	3.292-3.817	3.586-4.302	3.710-4.702	2.553-2.983
Sd error	0.272	0.079	0.133	0.181	0.257	0.086
Model 2	100 clusters					
Mean	0.981	0.672	0.781	0.901	0.893	0.597
$\sigma_{individual}$	5.112	3.868	4.120	4.960	4.487	3.173
Sd c.i.	4.408-5.900	3.623-4.128	3.728-4.530	4.459-5.508	3.910-5.123	2.932-3.442
Sd error	0.394	0.131	0.205	0.268	0.311	0.131
Mean	0.702	0.620	0.713	0.556	0.316	0.534
$\sigma_{individual:day}$	2.886	2.826	3.075	2.291	1.172	2.313
Sd c.i.	2.206-3.590	2.629-3.039	2.714-3.456	1.877-2.703	0.121-2.054	2.075-2.542
Sd error	0.353	0.105	0.194	0.210	0.498	0.101
Model 3	200 clusters					
Mean	0.774	0.451	0.545	0.676	0.745	0.455
$\sigma_{individual}$	4.330	2.882	3.400	3.833	3.838	2.743
Sd c.i.	3.798-4.888	2.734-3.037	3.157-3.660	3.505-4.188	3.370-4.344	2.569-2.921
Sd error	0.281	0.080	0.129	0.174	0.247	0.091
Model 4	200 clusters					
Mean	1.014	0.659	0.721	0.873	0.783	0.598
$\sigma_{individual}$	5.169	3.771	3.745	4.710	3.802	3.167
Sd c.i.	4.390-6.039	3.528-4.021	3.396-4.122	4.248-5.217	3.277-4.355	2.930-3.425
Sd error	0.418	0.125	0.187	0.248	0.281	0.128
Mean	0.763	0.620	0.666	0.551	0.172	0.538
$\sigma_{individual:day}$	3.146	2.826	2.875	2.271	0.576	2.327
Sd c.i.	2.417-3.940	2.616-3.039	2.527-3.219	1.877-2.676	0.022-1.395	2.100-2.562
Sd error	0.380	0.106	0.177	0.204	0.383	0.119

$$precision = \frac{TP}{TP + FP} \tag{9}$$

$$Kappa = \frac{\text{observed accuracy} - \text{expected accuracy}}{1 - \text{expected accuracy}} \tag{10}$$

The results indicate that accounting for the random effects results in significantly better overall predictive performance when compared to the boosted trees only model. The most complex models, the three-level models (models 2 and 4), show the best performance, both on the overall level, as well as for most transport modes. Using $k = 200$ (model 4) in the k-means algorithm, slightly increases performance compared to $k = 100$ (model 2). When we compare the performance for the separate modes for model 1 and 3 however, the mixed effects models do not always outperform the plain boosted trees model. Both balanced accuracy and the F1-score for bicycle are lower for mixed effect models for both model 1 and 3 compared to boosted trees only model, and for model 3 the F1-score for BTM is significantly lower for the mixed model compared to the boosted trees only score. Regarding Cohen's Kappa, all models fit the data very well. According to the classification by Landis and Koch (1977), values ranging from 0.61 to 0.80 are deemed substantial. Additionally, alternative perspectives categorize Kappa values between 0.40 and 0.75 as fair to good, and those exceeding 0.75 as excellent (Fleiss, 1981). Based on the commonly used Bayesian information criteration WAIC (Widely Applicable Information Criterion) (Vehtari et al., 2017; Watanabe, 2010), model 4 has the best fit. This would suggest that increasing the number of clusters in the k-means clustering step in the proposed algorithm, can lead to better model fits. Table 5 shows the results from previous work (Labee et al., In press.), where a similar algorithm was applied to estimate panel effects in decision trees using the CHAID tree algorithm. The algorithm in the cited work was slightly different, in the sense that after the fitting of the decision tree, indicators for the leaf nodes were used as fixed effects, rather than an indicator for the k-clusters as proposed in this work. Moreover, in Labee et al. (In press.) PT was still the merged choice of BTM & Train. The first model, CHAID 1, shows the best performance, both for the CHAID tree only as well as the mixed effects model. In this case, a two-level model was used to account for panel effects in each individual. Even though we can see a jump in performance once we account for the panel effects, the current algorithm proposed in this work using the boosted trees, shows a significantly larger jump, also compared to the CHAID mixed effects models.

Table 4: Predictive performance

Model	Mode	Boosted trees only				Mixed effects model				
		Bal. acc.	F1-score	Kappa	Overall acc.	Bal. acc.	F1-score	Kappa	Overall acc.	WAIC
				0.732	0.802			0.739	0.806	28,444.9
1	Bicycle	0.862	0.783			0.849	0.768			
	BTM	0.749	0.609			0.771	0.608			
	CaD	0.867	0.836			0.880	0.852			
	CaP	0.823	0.703			0.832	0.707			
	e-bike	0.823	0.727			0.833	0.725			
	Train	0.847	0.741			0.865	0.787			
	Walking	0.903	0.842			0.907	0.843			
								0.806	0.855	21,433.4
2	Bicycle					0.892	0.835			
	BTM					0.803	0.667			
	CaD					0.915	0.895			
	CaP		See model 1			0.877	0.795			
	e-bike					0.876	0.798			
	Train					0.868	0.790			
	Walking					0.927	0.868			
								0.743	0.809	27,885.2
3	Bicycle					0.857	0.780			
	BTM					0.727	0.523			
	CaD					0.879	0.851			
	CaP		See model 1			0.844	0.729			
	e-bike					0.828	0.716			
	Train					0.856	0.773			
	Walking					0.909	0.846			
								0.809	0.858	21,039.1
4	Bicycle					0.899	0.844			
	BTM					0.787	0.635			
	CaD					0.915	0.895			
	CaP		See model 1			0.884	0.802			
	e-bike					0.865	0.789			
	Train					0.862	0.778			
	Walking					0.927	0.876			

4 CONCLUSIONS AND FURTHER WORK

The primary objective of this study is to broaden the range of choices considered in travel mode choice, representing a crucial step toward extending the modelling of mode choice to each individual trip leg, and as such, multi-modal trips. In pursuit of this objective, we introduce a novel method designed to estimate random effects, while harnessing the predictive power of the boosted trees algorithm. We test this method to predict the travel mode for the main trip. Our analysis incorporates an expanded choice set encompassing seven distinct modes: bus/tram/metro, bicycle, car as a driver, car as a passenger, e-bike, train, and walking. This broader selection contrasts with the conventional inclusion of four to five modes in similar studies. The proposed algorithm, employing a single iteration, is tested through four distinct model specifications. These specifications are then benchmarked against a plain boosted trees model and CHAID decision tree. All of the proposed models outperform the plain boosted trees model, however, the proposed three-level models (model 2 and 4) show the best performance. It is noteworthy that Model 4 (the three-level model with 200 clusters) exhibits the best fit with the data based on the WAIC values. The estimates of parameters and the predictive performance metrics collectively emphasize the critical importance of accounting for random effects nested within different respondents when considering mode choice. This underscores the influence of individual-specific factors on the main trip mode choice. While the work addresses the main objective, the presented study is not without limitations. Future work should explore and compare alternative train/test partition strategies where an individual with its observations, either only occurs in the training data, or only in the testing data. Little knowledge is available both within the travel behavior community, as well as the general sense, on the possible 'data leakage' issue occurring when a conventional stratified split is used.

Table 5: Predictive performance CHAID tree

Model	Mode	CHAID tree only				Mixed effects model				
		Bal. acc.	F1-score	Kappa	Overall acc.	Bal. acc.	F1-score	Kappa	Overall acc.	WAIC
				0.656	0.746			0.713	0.785	24,958.8
CHAID 1[a]	Bicycle	0.809	0.702			0.816	0.726			
	CaD	0.831	0.794			0.871	0.844			
	CaP	0.777	0.618			0.842	0.709			
	e-bike	0.783	0.609			0.852	0.704			
	PT	0.772	0.634			0.851	0.658			
	Walking	0.887	0.816			0.890	0.821			
				0.502	0.639			0.673	0.755	28,897.3
CHAID 2[b]	Bicycle	0.684	0.510			0.805	0.695			
	CaD	0.769	0.726			0.844	0.811			
	CaP	0.695	0.494			0.832	0.699			
	e-bike	0.648	0.384			0.816	0.652			
	PT	0.655	0.444			0.828	0.657			
	Walking	0.853	0.723			0.872	0.793			
				0.493	0.634			0.659	0.745	29,179.3
CHAID 3[c]	Bicycle	0.679	0.502			0.800	0.683			
	CaD	0.761	0.718			0.838	0.811			
	CaP	0.699	0.500			0.817	0.699			
	e-bike	0.653	0.394			0.816	0.652			
	PT	0.644	0.412			0.823	0.657			
	Walking	0.847	0.721			0.863	0.793			
				0.456	0.611			0.665	0.749	29,413.3
CHAID 4[d]	Bicycle	0.696	0.518			0.797	0.683			
	CaD	0.746	0.701			0.839	0.805			
	CaP	0.664	0.436			0.833	0.696			
	e-bike	0.521	0.084			0.810	0.637			
	PT	0.613	0.351			0.816	0.635			
	Walking	0.835	0.708			0.874	0.794			

[a] level of significance merging: 0.05, level of significance splitting: 0.05, min. split: 20, min. node size: 7, max. height: infinite, its: 1
[b] level of significance merging: 0.05, level of significance splitting: 0.05, min. split: 100, min. node size: 50, max. height: 6, its: 1
[c] level of significance merging: 0.05, level of significance splitting: 0.01, min. split: 200, min. node size: 100, max. height: 6, its: 1
[d] level of significance merging: 0.05, level of significance splitting: 0.01, min. split: 200, min. node size: 100, max. height: 6, its: 100

ACKNOWLEDGEMENTS

This publication makes use of data from the Netherlands Mobility Panel, which is administered by KiM Netherlands Institute for Transport Policy Analysis. This work was funded by the project NEON, through the Dutch Research Council (NWO) Crossover Programme (project number 17628).

REFERENCES

Alan Agresti. Logit Models for Multinomial Responses. In *Categorical Data Analysis*, chapter 7, pages 267–313. John Wiley & Sons, Ltd, 2002. ISBN 978-0-471-24968-9. doi: 10.1002/0471249688.ch7.

Lars Böcker, Martin Dijst, and Jan Prillwitz. Impact of Everyday Weather on Individual Daily Travel Behaviours in Perspective: A Literature Review. *Transport Reviews*, 33(1):71–91, January 2013. ISSN 0144-1647, 1464-5327. doi: 10.1080/01441647.2012.747114.

Lars Böcker, Martin Dijst, and Jan Faber. Weather, transport mode choices and emotional travel experiences. *Transportation Research Part A: Policy and Practice*, 94:360–373, December 2016. ISSN 0965-8564. doi: 10.1016/j.tra.2016.09.021.

Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, October 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324.

Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. *Classification And Regression Trees*. Routledge, New York, 1984. ISBN 978-1-315-13947-0. doi: 10.1201/9781315139470.

Paul-Christian Bürkner. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28, 2017. doi: 10.18637/jss.v080.i01.

Robert Cervero. Built environments and mode choice: Toward a normative framework. *Transportation Research Part D: Transport and Environment*, 7(4):265–284, June 2002. ISSN 13619209. doi: 10.1016/S1361-9209(01)00024-4.

Robert Cervero and Kara Kockelman. Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 2(3):199–219, September 1997. ISSN 13619209. doi: 10.1016/S1361-9209(97)00009-6.

Long Cheng, Xuewu Chen, Jonas De Vos, Xinjun Lai, and Frank Witlox. Applying a random forest method approach to model travel mode choice behavior. *Travel Behaviour and Society*, 14:1–10, January 2019. ISSN 2214367X. doi: 10.1016/j.tbs.2018.09.002.

Reid Ewing and Robert Cervero. Travel and the Built Environment: A Meta-Analysis. *Journal of the American Planning Association*, 76(3):265–294, June 2010. ISSN 0194-4363, 1939-0130. doi: 10.1080/01944361003766766.

J. L. Fleiss. *Statistical Methods for Rates and Proportions. 2nd Ed.* Wiley Series in Probability and Mathematical Statistics. Wiley, New York, 1981. ISBN 978-0-471-06428-2.

M. Fokkema, N. Smits, A. Zeileis, T. Hothorn, and H. Kelderman. Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior Research Methods*, 50(5):2016–2034, October 2018. ISSN 1554-3528. doi: 10.3758/s13428-017-0971-x.

Marjolein Fokkema, Julian Edbrooke-Childs, and Miranda Wolpert. Generalized linear mixed-model (GLMM) trees: A flexible decision-tree method for multilevel and longitudinal data. *Psychotherapy Research*, 31(3):329–341, April 2021. ISSN 1050-3307. doi: 10.1080/10503307.2020.1785037.

Julian Hagenauer and Marco Helbich. A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications*, 78:273–282, July 2017. ISSN 09574174. doi: 10.1016/j.eswa.2017.01.057.

Ahlem Hajjem, François Bellavance, and Denis Larocque. Mixed effects regression trees for clustered data. *Statistics & Probability Letters*, 81(4):451–459, April 2011. ISSN 0167-7152. doi: 10.1016/j.spl.2010.12.003.

Ahlem Hajjem, François Bellavance, and Denis Larocque. Mixed-effects random forest for clustered data. *Journal of Statistical Computation and Simulation*, 84(6):1313–1328, June 2014. ISSN 0094-9655, 1563-5163. doi: 10.1080/00949655.2012.741599.

Ahlem Hajjem, Denis Larocque, and François Bellavance. Generalized mixed effects regression trees. *Statistics & Probability Letters*, 126:114–118, July 2017. ISSN 0167-7152. doi: 10.1016/j.spl.2017.02.033.

Sascha Hoogendoorn-Lanser, Nina T.W. Schaap, and Marie-José OldeKalter. The Netherlands Mobility Panel: An Innovative Design Approach for Web-based Longitudinal Travel Data Collection. *Transportation Research Procedia*, 11:311–329, 2015. ISSN 23521465. doi: 10.1016/j.trpro.2015.12.027.

Michael Hyland, Charlotte Frei, Andreas Frei, and Hani S. Mahmassani. Riders on the storm: Exploring weather and seasonality effects on commute mode choice in Chicago. *Travel Behaviour and Society*, 13: 44–60, October 2018. ISSN 2214367X. doi: 10.1016/j.tbs.2018.05.001.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Kyoungok Kim. Effects of Weather and Calendar Events on Mode-Choice Behaviors for Public Transportation. *Journal of Transportation Engineering, Part A: Systems*, 146(7):04020056, July 2020. ISSN 2473-2907, 2473-2893. doi: 10.1061/JTEPBS.0000371.

Kyusik Kim, Kyusang Kwon, and Mark W. Horner. Examining the Effects of the Built Environment on Travel Mode Choice across Different Age Groups in Seoul using a Random Forest Method. *Transportation Research Record*, 2675(8):670–683, August 2021. ISSN 0361-1981. doi: 10.1177/03611981211000750.

Seheon Kim, Soora Rasouli, Harry Timmermans, and Dujuan Yang. Estimating panel effects in probabilistic representations of dynamic decision trees using bayesian generalized linear mixture models. *Transportation Research Part B: Methodological*, 111:168–184, May 2018. ISSN 0191-2615. doi: 10.1016/j.trb.2018.03.010.

Pim Labee, Soora Rasouli, and Seheon Kim. Predicting mode choice using decision trees while accounting for random effects using panel data. In *17th International Conference on Travel Behavior Research, 14-18 July, 2024*, page X, Vienna, Austria, In press.

Nan M. Laird and James H. Ware. Random-Effects Models for Longitudinal Data. *Biometrics*, 38(4): 963–974, 1982. ISSN 0006-341X. doi: 10.2307/2529876.

J. Richard Landis and Gary G. Koch. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174, 1977. ISSN 0006-341X. doi: 10.2307/2529310.

Michael D Lee and Eric-Jan Wagenmakers. Bayesian Cognitive Modeling: A Practical Course. 2013.

Chengxi Liu, Yusak O. Susilo, and Anders Karlström. The influence of weather characteristics variability on individual's travel mode choice in different seasons and regions in Sweden. *Transport Policy*, 41: 147–158, July 2015. ISSN 0967070X. doi: 10.1016/j.tranpol.2015.01.001.

Chengxi Liu, Yusak O. Susilo, and Anders Karlström. Weather variability and travel behaviour – what we know and what we do not know. *Transport Reviews*, 37(6):715–741, November 2017. ISSN 0144-1647, 1464-5327. doi: 10.1080/01441647.2017.1293188.

Scott M Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, volume 5.1, pages 281–298. University of California Press, January 1967.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2022.

Rebecca J. Sela and Jeffrey S. Simonoff. RE-EM trees: A data mining approach for longitudinal and clustered data. *Machine Learning*, 86(2):169–207, February 2012. ISSN 1573-0565. doi: 10.1007/s10994-011-5258-3.

Yu Shi, Guolin Ke, Damien Soukhavong, James Lamb, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu, and Nikita Titov. *Lightgbm: Light Gradient Boosting Machine*, 2023.

Yan Song, Louis Merlin, and Daniel Rodriguez. Comparing measures of urban land use mix. *Computers, Environment and Urban Systems*, 42:1–13, November 2013. ISSN 0198-9715. doi: 10.1016/j.compenvurbsys.2013.08.001.

Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432, September 2017. ISSN 1573-1375. doi: 10.1007/s11222-016-9696-4.

Sumio Watanabe. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research*, 11(116):3571–3594, 2010. ISSN 1533-7928.

Xilei Zhao, Xiang Yan, Alan Yu, and Pascal Van Hentenryck. Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel Behaviour and Society*, 20: 22–35, July 2020. ISSN 2214-367X. doi: 10.1016/j.tbs.2020.02.003.