A machine learning approach to mode choice analysis

BENV0095: Energy and Transport Analytics

HQCJ5 22/05/2020

Total Words :	2992
Headers :	16
Math Inline :	50
Math Display :	7

LaTex word count

Contents

List of Figures

List of Tables

I Introduction

Our desire to gravitate toward, and conduct business in the globes thriving economic centres has created a number of detrimental consequences. One of the most damaging is air pollution, responsible for 40,000 deaths in the UK each year (Physicians 2019). Recent analysis of schools in the Breathe London sensor network (London 2020), show that all schools where children are exposed to illegal No2 levels of pollution (40 µg/m^3) were in the city centre, in boroughs with either the highest levels of economic activity or vehicle congestion.

Figure 1: London schools in Breathe London network with illegal levels of pollution. [1] *Data source: Breathe London*

As shown in [2], recent analysis of the sensor network during the Covid-19 pandemic highlights the link between air pollution and traffic, where No2 levels dropped between 20 and 24 % between 13th March and 13th April in the ULEZ. Clearly, changes in transport behaviour can have positive impacts on urban life. These impacts aren't limited to air pollution, but also include reduced noise pollution, lower commuting times and a greater availability of space in cities for alternative activities.

Policies can and should be implemented to promote positive social practices with respect to transport which accelerate behavioural change and adjust social norms. As shown in [3], transport related behaviours have changed over a ten year period. A less trivial exercise is to understand why. Indeed, if policies are to be designed to promote positive transport behaviours, then the drivers between individual choice and preference need to be understood. Understanding these drivers may allow for targeted information dissemination regarding alternatives, better infrastructure planning and implementation and smarter normative feedback mechanisms.

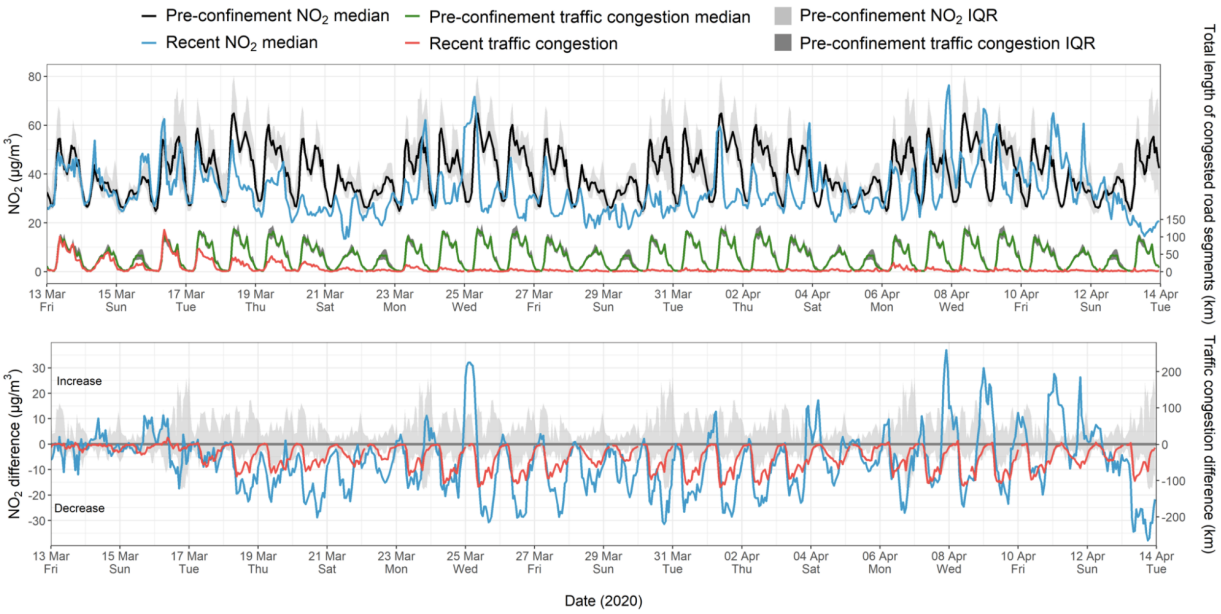

Figure 2: A comparison of median congestion and pollution levels in the ULEZ. [2] *Data source: Breathe London and Waze*

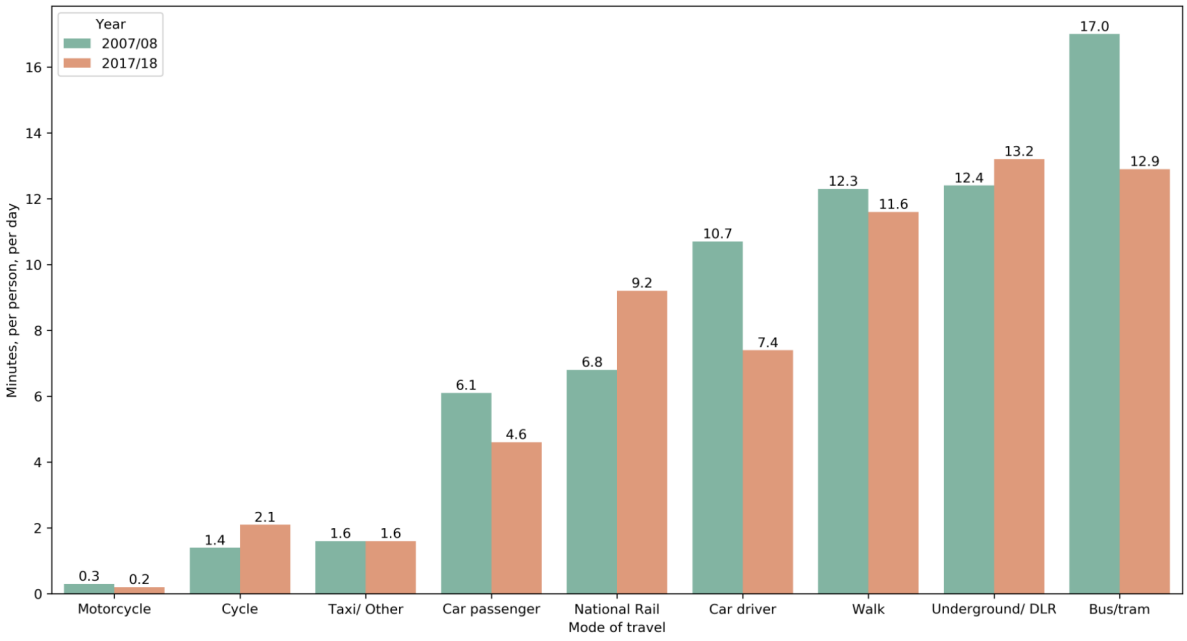

Figure 3: Changes in London transport mode utilisation over a ten year period. [3] *Data source: London Travel Demand Survey (LTDS)*

Travel mode surveys are a well established method in collecting data on individual travel choices. Interpreting these data typically involves methods between the fields economics, social science and statistics, with utility maximisation theory taking a leading role. Recently, literature has indicated machine learning could be highly applicable in predicting and understanding user mode choice. Machine learning is often cited as a 'black box', where the relationships between inputs and output are complex and unknowable (Breiman 2001). If true, this would imply machine learning is unsuitable in explaining complex relationships which constitute human behaviour. This paper aims to show that machine learning is a valid approach in explaining mode choice behaviour, and that highly complex, non-linear models can still be interpreted in a meaningful way.

II Literature review

Mcfadden (Mcfadden 1974) first introduced the discrete choice model as a method to understand influential factors in mode choice, with the Multinomial Logit Model (MNL) becoming the cornerstone of mode choice analysis. (Chen, Liu, and Li 2013) highlights that the independence of irrelevant alternatives (IIA) property of MNL models is a major constraint and since Mcfadden, a number of statistical alternatives have been proposed to relax IIA including the nested logit (NL) (Dubin and McFadden 1984) and the mixed multinomial logit model (MMNL) (Sheffi, Hall, and Daganzo 1982). (Chang 2019) highlights that discrete choice models are often analysed statically, which limits their applicability in long term forecasting, and cites a number of alternatives including Markov-Chain methods and machine learning. Whilst Chang et al. claims that machine learning methods have more "flexible structures than traditional logit approaches to represent the relationship between the attributes of alternatives and choices", the paper only analyses the performance of models in terms of validation loss. Indeed, most leading literature on the topic follow the same theme, where model performance takes a leading role rather that interoperability and feature analysis. (S. Wang 2019), benchmark the performance of 86 ML models and identified gradient boosting as the most accurate, with lowest variance in predictions. This in part, motivates the use of a gradient boosting machine (GMN) in this study. (F. Wang and Ross 2018) Compare the performance of MNL models and GBM, but pay little attention to feature importance or interactions. This paper aims to begin to fill the gap between the worlds of discrete choice theorists and machine learning, and instead of aiming to optimise model performance, highlight the appropriateness of the method and identify actionable insights from trained models.

III Methodology

III.I Data description and exploratory data analysis

This paper uses a subset of data from a travel mode survey conducted in Thessaloniki, Greece. The corresponding travel modes, or target variables were Walk, Car, Bus and Motorcycle[1]. Independent variables in the raw dataset are listed below. [2]
The raw data required significant processing. It was found that a number of start and end time entries were recorded in reverse, this was handled by taking the absolute time differences between these features and recalculating trip duration. Since trip time holds so statistical relationship with the cyclical nature of time, two additional features were engineered - sin time and cosine time, the original time variables were then dropped.

Two datasets were created to handle categorical variables differently. Discrete choice logit models were trained using a dataset of one hot encoded categories, where numerical features were zero mean scaled. A constant feature was also created in order to estimate intercept values. For tree based predictors, categorical variables were label encoded. Missing values in the "income" feature were encoded as "-1", meaning these values could be removed in one split, reducing tree depth. [3] As shown in [4], spearman rank order correlations were calculated, to facilitate feature selection and importance analysis in the following sections.

[1] Due to significant class imbalance, Motocycle class was removed.

[2] variables with no significance to target variables were removed, including "trip id" and "questionnaire number".

[3] A model could have been trained to predict these missing values, however, this was not conducted to avoid adding bias to the data.

Variable name	Variable description
Income	Declared Income
Gender	Declared Gender
Age	Declared Age
Occupation	Declared occupation
Education	Declared education level
HouseHoldSize	Total household size
HouseHold16	Number of household members over the age of 16
driversLicense	Boolean (1=yes, 0=no)
carAvail	Boolean (1=yes, 0=no)
OriginLoc	Municipality of trip origin
DestLoc	Municipality of trip destination
startActiv	Activity performed at location of origin
endActiv	Activity performed at location of destination
tripStart time	Declared trip start time
tripEnd time	Declared trip end time
tripDuration	Trip duration
TripDay	Day of the week the trip was made
realTravelTime.Walk	Google travel time - (seconds)
realShortestDistance.Walk	Google travel distance - (meters)
realTravelTime.Car	Google travel time - (seconds)
realShortestDistance.Car	Google travel distance - (meters)
realTravelTime.Bus	Google Travel time - (seconds)
realShortestDistance.Bus	Google travel distance - (meters)
realTransfer.Bus	Number of bus transfers
realAccessWalkTime.Bus	Access to bus stop walk travel time
realAccessWalkDist.Bus	Access to bus stop walk distance
realEgressWalkTime.Bus	Egress to destination walk time
realEgressWalkDist.Bus	Egress to destination distance

Table 1: Independent variables used in study from Greek travel survey [1]

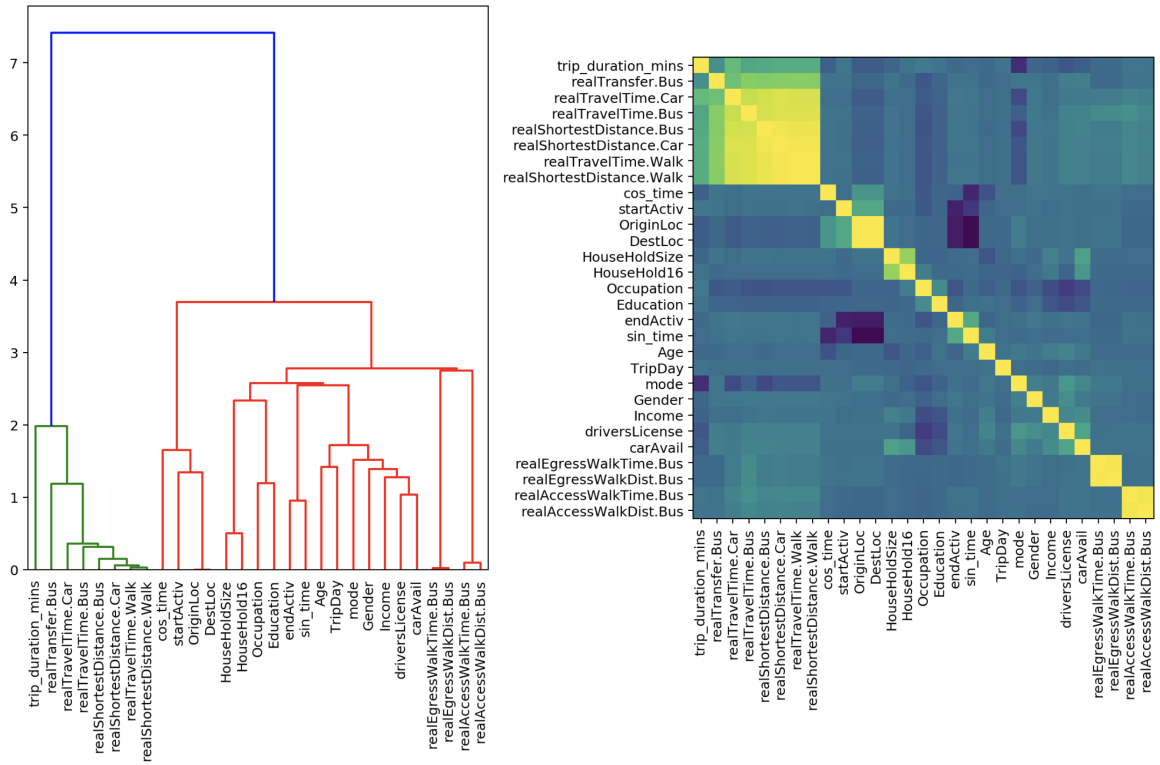

Figure 4: Spearman rank order matrix and clustering by dendrogram for independent variables (yellow=high correlation, dark blue=low correlation) [4]

III.II Method

The first analysis approach conducted was a logistic discrete choice model. Model coefficients were estimated using Newton Raphson optimisation. Prediction accuracy of this model was also identified using a standard SKlearn implementation of multi-class logistic regression in python. The paper then uses more sophisticated tree based predictors to enhance predictive accrurracy. Multiple feature importance methods are considered, followed by calculating Shapley Additive Explanation values (SHAP) for a gradient boosting model. The code behind this analysis can be found on the following github page.

Github link: https://github.com/Olliejp/mode-choice-analysis

IV Discrete choice analysis

Discrete choice theory is grounded in utility functions, whereby an individual is assumed to behave in a way which maximises utility (Chen, Liu, and Li 2013). That is to say there is somewhat a linear relationship between behavioural change or decision making and the utility of an alternative. The theory can be expressed as follows.

$$\mathbf{U}_{nj} = \mathbf{V}_{nj} + \epsilon_{nj} \tag{1}$$

Where $\mathbf{U}_{nj}$ represents the utility that the individual n associates with alternative j, $\mathbf{V}_{nj}$ represents measurable utility associated with an alternative and ϵ_{nj} is the error term (or unmeasurable utility). Given the assumption of rationality from utility maximisation theory, we can therefore simply express the likelihood of travel mode i being chosen as that which represents the greatest probability given a set of alternatives, where.

$$\mathbf{P}_{ni} = \mathbf{P}(\mathbf{V}_{ni} + \epsilon_{ni} > \mathbf{V}_{nj} + \epsilon_{nj}, \forall j \neq i), i, j \in \mathbf{J} \tag{2}$$

Given one wishes to calculate the probability of a single alternative, and given the assumption that the error term is IID, (2) takes the following form in a multinomial case (MNL), where $\mathbf{V}_{nj}$ is $\in$ of a set of $\mathbf{J}$ alternatives.

$$\mathbf{P}_{ni} = \frac{exp^{\mathbf{V}_{ni}}}{\sum_{j=1}^{\mathbf{J}} exp^{\mathbf{V}_{nj}}} \tag{3}$$

In a binary base the probability of alternative $\mathbf{P} = 1$ or $1 - \mathbf{P} \neq 1$ is as (4).

$$\mathbf{P}_{ni} = \frac{exp^{\mathbf{V}_{ni}}}{1 + exp^{\mathbf{V}_{ni}}} \tag{4}$$

IV.I Weight analysis travel mode survey in a MNL context

Using the Bus travel mode as a base case the weights which constitute the utility associated with choosing the Car travel mode were estimated using Newtons optimisation method. The twelve most significant features are displayed in *table 2*.

Variable name	**Variable code**	**Weight**	**P-Value**
HouseHold16	HH16	-0.2540	< 0.001
driversLicense	DL	2.0443	< 0.001
carAvail	CA	2.2942	< 0.001
realTravelTime.Car	RTC	-0.0023	< 0.001
realTravelTime.Bus	RTB	0.0008	< 0.001
realTransfer.Bus	TB	0.1629	0.061
realAccessWalkDist.Bus	RAWDB	-7.736e-05	0.011
realAccessWalkTime.Bus	RAWTB	0.0003	0.053
trip.duration.mins	TD	-0.0412	< 0.001
sin.time	ST	-0.1168	0.033
cos.time	CT	0.3811	< 0.001
realShortestDistance.Bus	TSDB	0.0001	< 0.001
Intercept	I	-2.1102	< 0.001

Table 2: Weights and significance level for $\mathbf{P}_{car} = 1$ [2]

Using features with p-values < 0.05 we can define a linear expression for the utility of travel mode Car for a given individual n.

$$\begin{aligned}\mathbf{V}_{n,Car} = &-2.1102 - 0.254(HH16) + 2.0443(DL) + 2.2942(CA) \\ &- 0.0023(RTC) + 0.0008(RTB) - 0.000077(RAWDB) - 0.0412(TD) \\ &- 0.1168(ST) + 0.3811(CT) + 0.0001(TSDB)\end{aligned} \tag{5}$$

By substituting (5) into (4) we can derive the probability that a given individual will choose mode Car over the base case of Bus. For classification purposes, a decision boundary would be optimised through cross validation.

IV.II Discrete choice method analysis

There are several merits to logit based mode choice analysis methods. Firstly, weight interpretation is trivial and usually intuitive. From *table 2* we observe that having a drivers licence increases the log odds of choosing mode Car by approximately a factor of 2. Similarly having access to a car increases log odds by a factor of 2.29. These are easily understandable results.

We observe that there is significance in sin and cosine time, indicating that the time of day influences ones decision to choose a bus or car travel mode, possibly explained by safety or traffic concerns.

There are however, a number of shortcomings in this commonly accepted approach to mode choice analysis. The greatest is this methods grounding in the concept of utility. Whilst traditional economists support concepts of rational choice and utility theory, behavioural economics suggests that this is almost never consistent, were there are common biases and heuristics which cause humans to regularly divert from actions that would maximise our health, wealth and happiness (Gaker 2011). Expressing this notion more mathematically, human behaviour, and the function mapping between utility and choice may be highly non-linear. It therefore seems an oversimplification to express discrete choice with a logit model where there is a linear relationship between log odds and feature permutations. MNL models also assume utility random terms of alternatives parts have totally independent structures (Chen, Liu, and Li 2013), which, in reality will rarely be the case. There may also be strong interaction effects between features which become lost in MNL analysis. Finally, weight analysis and feature importance becomes meaningless if the underlying model is a weak interpretation of the true function mapping $x :\rightarrow y$. Great attention should therefore be paid to model performance and optimisation choice.

V Non-linear machine learning analysis

There are a number of factors motivating more sophisticated machine learning applications in mode choice analysis. Firstly, using tree ensembles, or deep neural networks with non-linear activation functions, non-linear hyperplane boundaries can be estimated and non-linear structure between features preserved. Tree ensembles and neural networks may also achieve higher validation accuracy than a single logit layer, meaning a more accurate estimation of the true function. Further, travel mode surveys may be large, particularly if on a national scale. The second order optimisation method conducted in the previous section involves an expensive hessian inversion equating to a time complexity of $O(n^3)$. A decision tree, by comparison can be built in $n(logn)$ time if features are pre-sorted. Whilst machine learning may have a number of merits, literature still claims interoperability as a major drawback. The following sections illustrate how machine learning could be applied in a travel mode analysis context whilst maintaining a high level of model explainability.

V.I Model performance analysis

The following table and figures highlight the predictive performance of the model estimated in section 4, a single decision tree with max depth of 3, and a GBM. Optimum parameters for the GBM were found using grid search and 5 fold cross-validation across 500 models [4].

Model	**F1 Accuracy**
Logistic Regression (Softmax activation)	0.83
Decision tree (max depth 3)	0.79
Gradient Boosting Machine	0.89

Table 3: Model accuracy comparison [3]

[4] Accuracy scores above 90% were achieved by bagging GBM models, and stacking. This procedure does reduce model interoperability. Since absolute accuracy is not within the scope of this paper, these results were omitted.

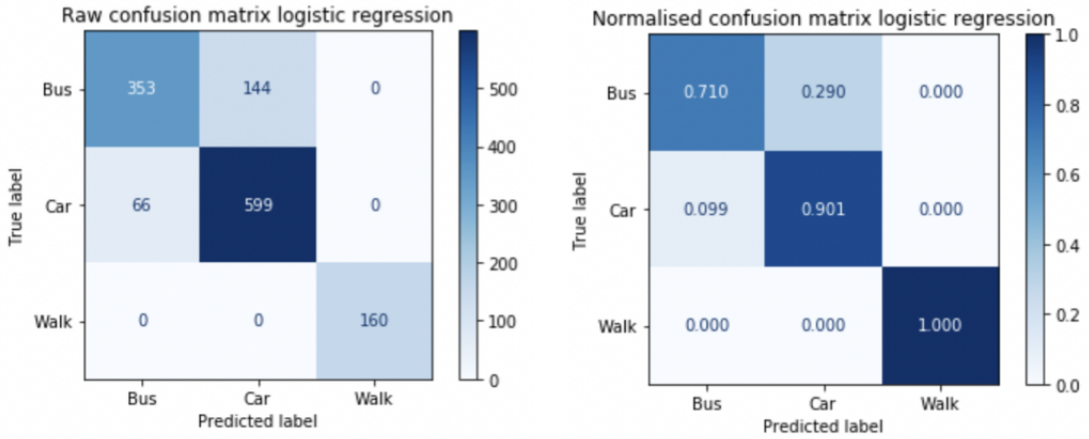

Figure 5: Confusion matrix for logistic regression [5]

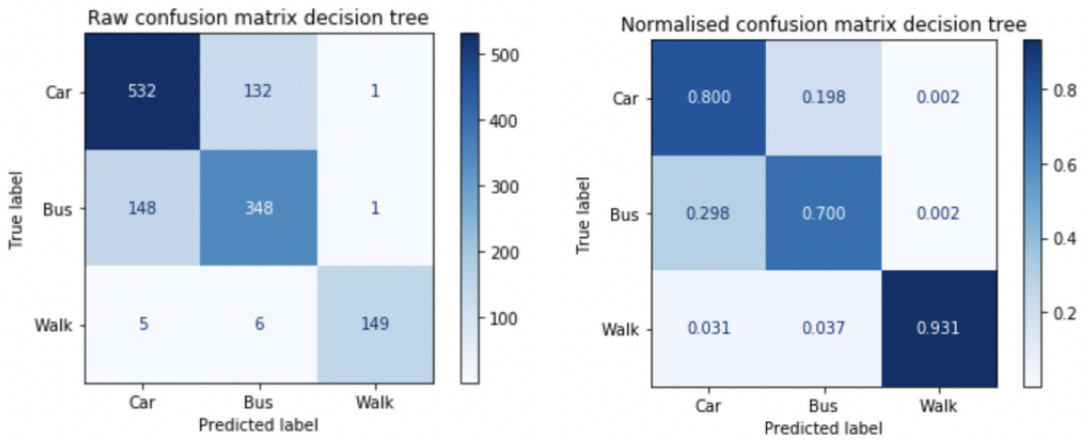

Figure 6: Confusion matrix for single tree [6]

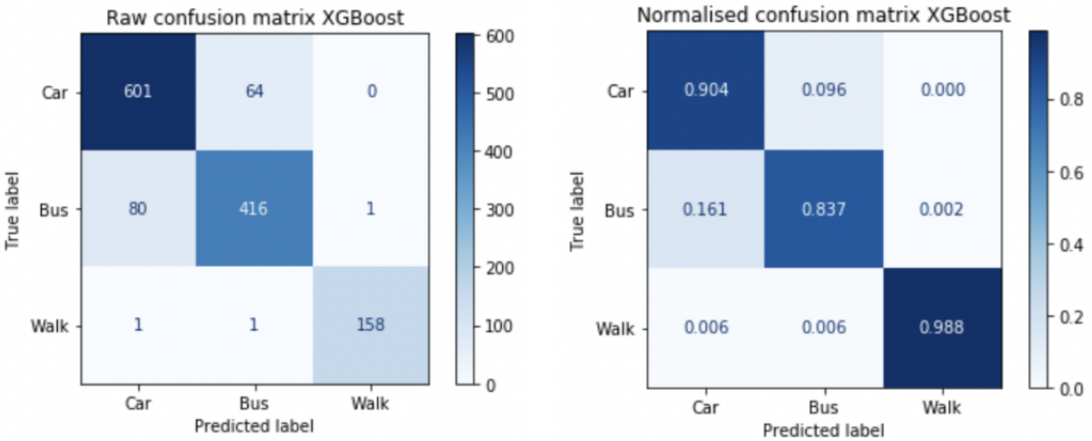

Figure 7: Confusion matrix for GBM [7]

The above results show that ensemble tree models can achieve a lower overall classification error than softmax regression. It is also noted that the minority walk class is highly separable. This, in fact can cause convergence issues for iterative solvers like gradient descent used in the softmax model. Accuracy was determined using a hold out validation set, using a random 80:20 train and validation split.

V.II Tree ensembles

Single, shallow trees are often weak predictors, as seen in the above model with a max depth of 3. Deep models, however tend to overfit on training data. The advantage of using trees, however, is that their structure easily explains how a particular prediction is made. [8] illustrates this statement.

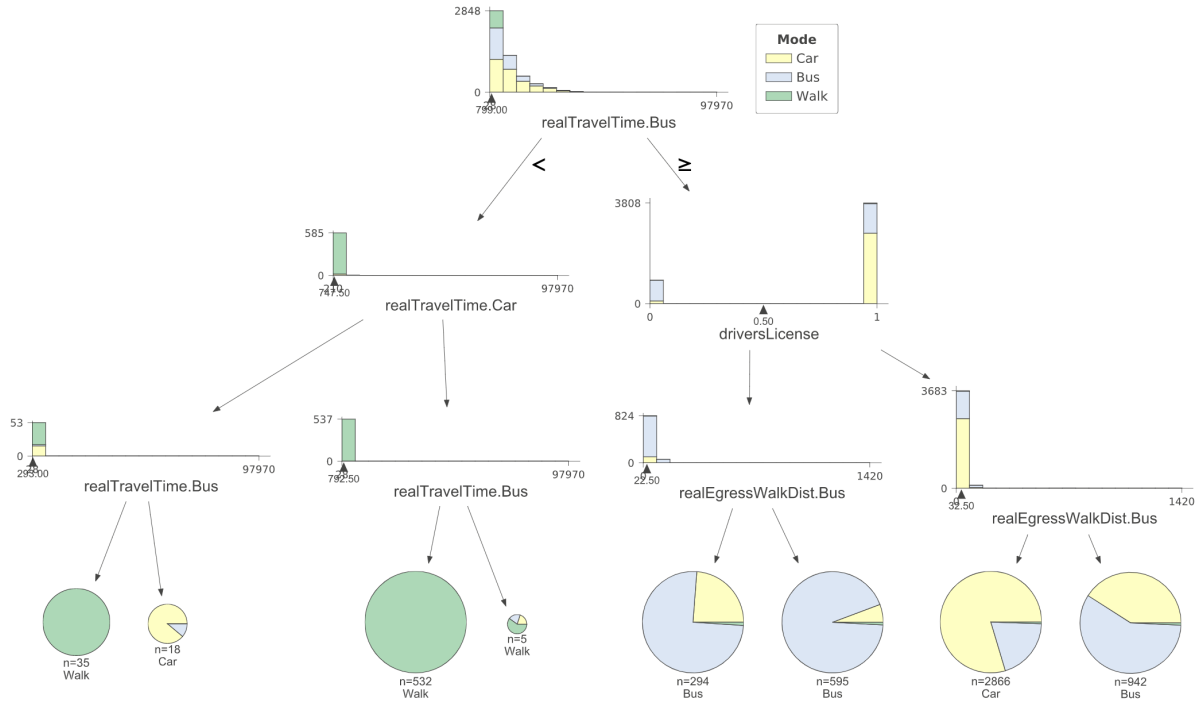

Figure 8: Visualisation of single tree splits, max depth 3 [8]

GBM's overcome this issue by fitting numerous shallow trees in a stage-wise fashion. For a single tree, the tree structure is directly optimised, however, since this is not a function that can be differentiated, GBM's instead optimise the residuals of the k previous trees. Each split is therefore given by the minimisation of (6)[5].

$$loss^{k+1} = \sum_{i=1}^{\mathbf{N}} (y_i - f_k(x_i) - f_{k+1}(x_i))^2 \tag{6}$$

Since we optimise the residuals and not the tree structure itself, we lose the direct interoperability provided my the single tree in [8], and must use alternative statistical methods.

[5]Where $y_i - f_k(x_i)$ are the pseudo residuals (residuals derived for difference between calculated probability and true class) from the previous tree.

V.III Global feature importance

A naive method commonly cited to identify feature importance in tree ensembles is total gain.

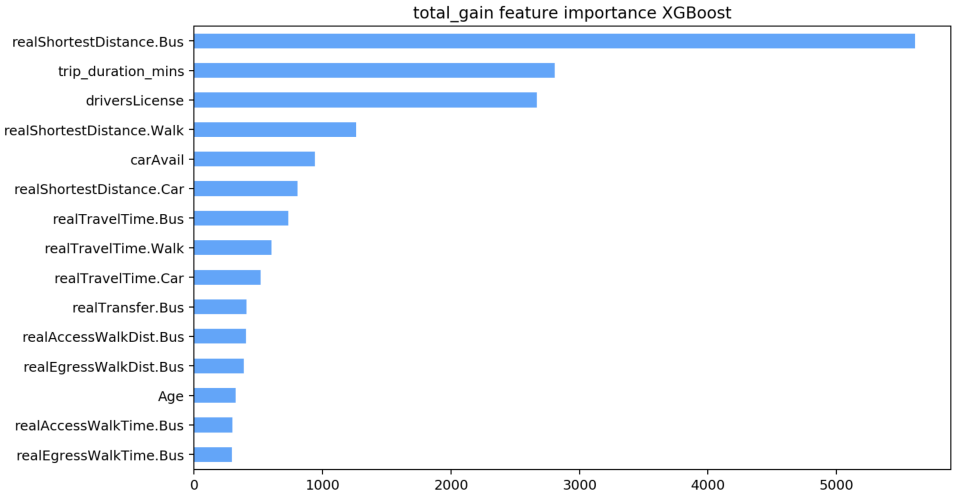

Figure 9: Total gain scores for GBM [9]

Total gain is the summation and average of gini decrease for each feature over each tree, and can often provide feature importance close to the ground truth. We observe a number of "highly important" features after calculating total gain is consistent with significant features found in our MNL model. This method can, however, be highly biased particularly where predictor variables vary in their scale of measurement or their number of categories, as is the case in our GMB trained dataset. A far more consistent method to measure global feature importance is permutation or drop one column importance (Parr 2016). This method however, has difficulties where there is colinearity amongst features, since correlated features need to be dropped together for accurate results. Given the high correlation amongst features clusters shown in [4], permutation importance was not calculated. Analysing global feature importance for complex ensembles is useful, since we have a picture what what factors are most influential in individual mode choice. A more useful exercise, however, is to understand local feature importance, for specific individuals, and understand the interaction effects between features.

V.IV Game theory - Shapley Additive Exlanations (SHAP)

For a linear model, as in (5), calculating how each feature effects a datapoint is trivial. The contribution of the $j'th$ feature ϕ_j, to the prediction $\hat{f}(x)$ is given by (7).

$$\phi_j(\hat{f}) = w_j x_j - \mathbf{E}(w_j \mathbf{X}_j) \tag{7}$$

Where w_j is the $j'th$ model weight and $\mathbf{E}(w_j \mathbf{X}_j)$ is the mean effect estimate for feature j. The $j'th$ contribution is therefore the difference between the feature effect and the mean effect. By repeating this process for all features, we can understand the positive and negative effect for each feature for a specific training example.

These effects are far more challenging to calculate for complex non-linear models such as the GMB used in this paper. We instead, turn to game theory.

SHAP values (Lundberg and Lee 2017) (Molnar 2019) explain a prediction by assuming each feature is a player in a game where the prediction is a "payout" distributed amongst players. The exact formulation is given by (8), for an $n * m$ data matrix the SHAP value of the $j'th$ feature is its contribution to the payout, weighted and summed over all possible feature combinations.

$$\phi_j(SHAP) = \sum_{S \subseteq \{x_1,...,x_m\} \setminus x_j} \frac{|S|!(m - |S| - 1)!}{m!}(SHAP(S \cup \{x_j\}) - SHAP(S)) \tag{8}$$

Where S is a subset from the original feature space, x is the vector of feature values of the example to be explained. $SHAP_x(S)$ is the prediction for features in set S, marginalised over features not in S. Calculating SHAP values for more than a few features is NP-hard, so instead the values are approximated using Monte-Carlo sampling through the following.

Result: SHAP value for i, j feature value
Iterations M, example x, feature index j, data matrix X, model f;
while *iteration* $<$ *len M* **do**
  Draw random example z from data matrix X;
  Choose random permutation o of feature values;
  Order example x: $x_o = (x_{(1)}, ..., x_{(j)}), ...x_{(m)}$;
  Order example z: $z_o = (z_{(1)}, ..., z_{(j)}), ...z_{(m)}$;
  Build two new examples;
    With feature j: $x_{+j} = (x_{(1)}, ..., x_{(j-1)}, x_{(j)}, z_{(j+1)}, ..., z_{(m)})$;
    Without feature j: $x_{-j} = (x_{(1)}, ..., x_{(j-1)}, z_{(j)}, z_{(j+1)}, ..., z_{(m)})$;
  Compute marginal contribution: $\phi_j^{marg} = \hat{f}(x_{+j}) - \hat{f}(x{-j})$;
  Compute SHAP value as the average: $\phi_j(x) = \frac{1}{M} \sum_{m=1}^{M} \phi_j^{marg}$;
end

Algorithm 1: SHAP estimation for a single feature value

V.V SHAP in practice for mode choice analysis using Gradient Boosting Machine

As a tensor of as many 3rd dimensions as class labels, we can separately plot the global feature importance as their positive and negative contribution to a class prediction. The following highlight global SHAP values for car and bus travel modes [6][7].

[6] Where classes are boolean, high feature value is 1, low is 0

[7] The possible scope of SHAP analysis on this dataset is too vast for a thorough investigation in this paper, more detailed analysis can be seen in the supporting Jupyter Notebook

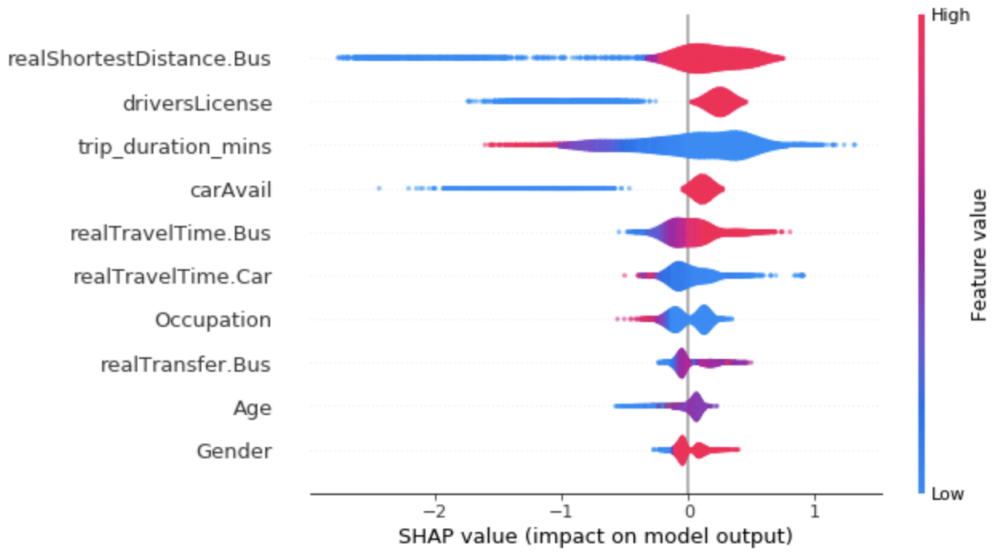

Figure 10: SHAP contributions for prediction Car travel mode for top 10 most important features for that class [10]

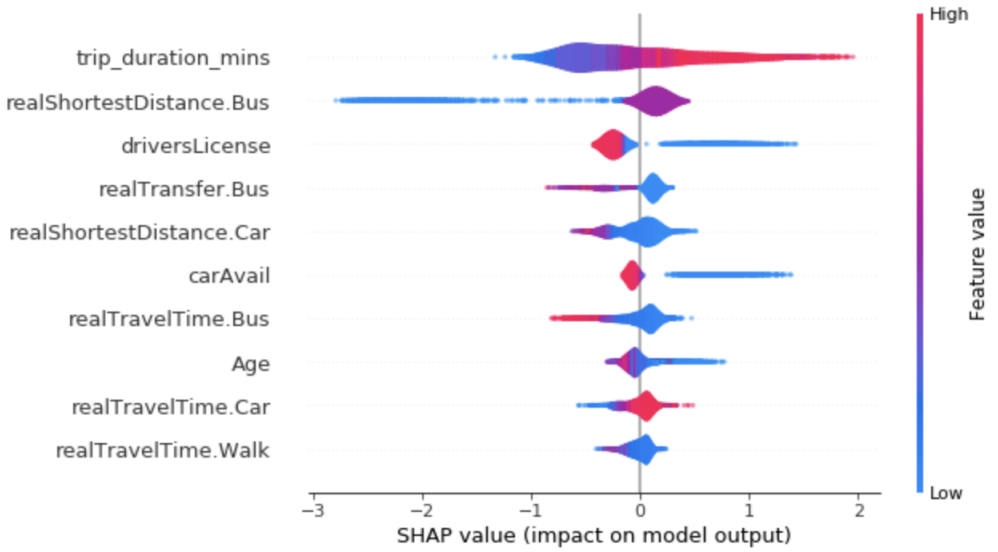

Figure 11: SHAP contributions for prediction Bus travel mode for top 10 most important features for that class [11]

Clearly the real distances and times between locations have a strong influence of whether a car or bus mode is taken. Intuitively having a drivers licence means the model is more likely to predict an individual takes a car. Age and gender also seem to have some influence on class prediction. Whilst a similar analysis could be made with a simpler MNL model, SHAP analysis allows us to investigate subtle non-linear feature interactions which would be lost with a linear model.

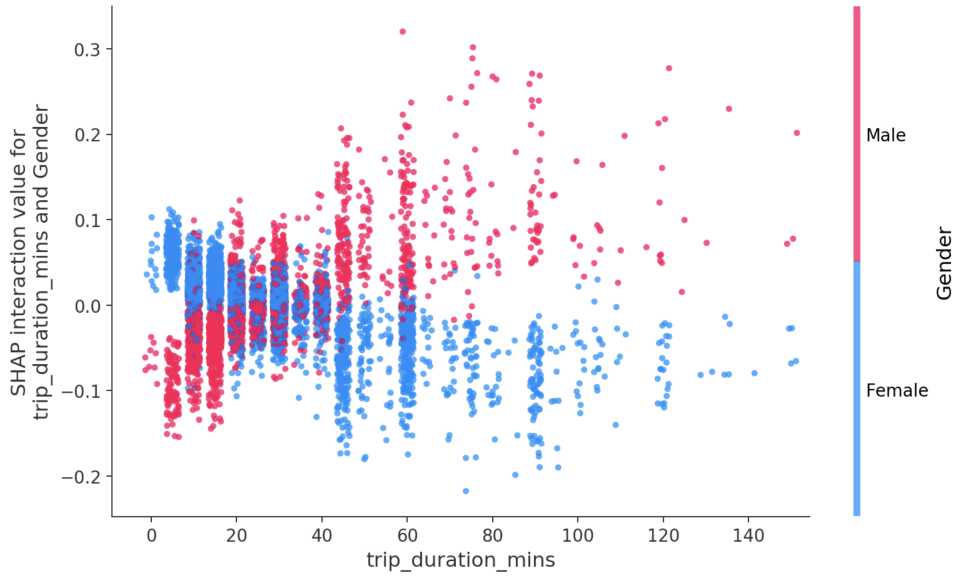

Figure 12: Interaction effects between travel time and gender for Car travel mode class [12]

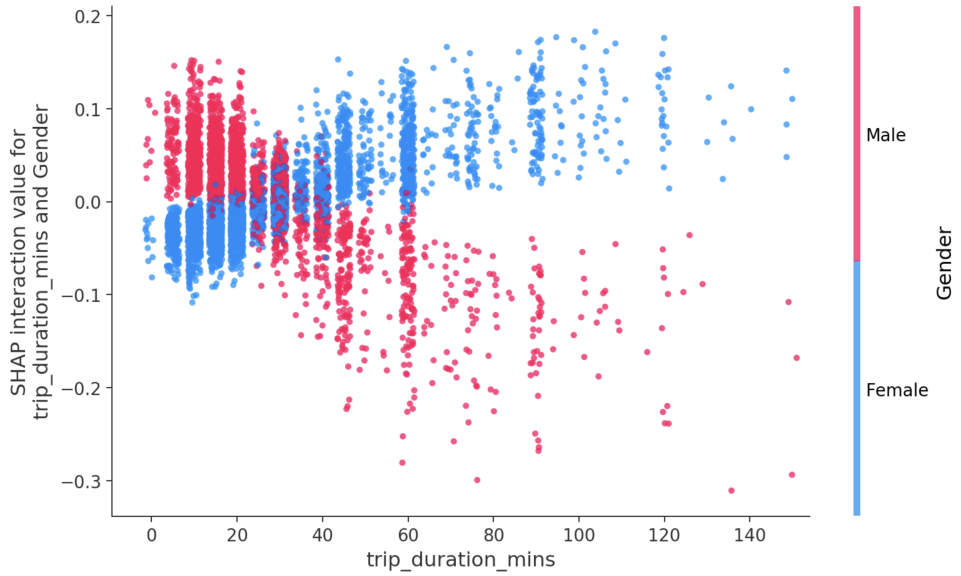

Figure 13: Interaction effects between travel time and gender for Bus travel mode class [13]

Analysis of this dataset shows that females are more likely to take a bus than males. [12][13] show an interesting non-linear interaction between trip duration and gender. With respect to prediction for Car mode, males are more likely to travel by car for longer trips. The opposite is true for the Bus mode, where females are more likely to take longer journeys by bus. This could be explained by safety concerns by females, or males being less prone to spending long duration's on public transport.

By indexing our SHAP matrix for a specific example, or person, we can understand what factors most strongly influence individual choice for a travel mode given specific variables. Randomly selecting a female from the validation set travelling from home to school with the following profile, [14][15][16] show how each variable influences the model prediction.

Variable	Value
Gender	Female
Age	16-24
Occupation	Student
Education	Lyceum
HouseHoldSize	2
driversLicense	1
startActiv	Home
endActiv	School
TripDay	Saturday
realTravelTime.Walk	1163
realShortestDistance.Walk	1565
realTravelTime.Car	355
realShortestDistance.Car	2026
realTravelTime.Bus	745
realShortestDistance.Bus	355
realTransfer.Bus	1
realAccessWalkTime.Bus	242
realAccessWalkDist.Bus	359
trip.duration.mins	15

Table 4: Feature values for a female who chose travel mode Walk [4]

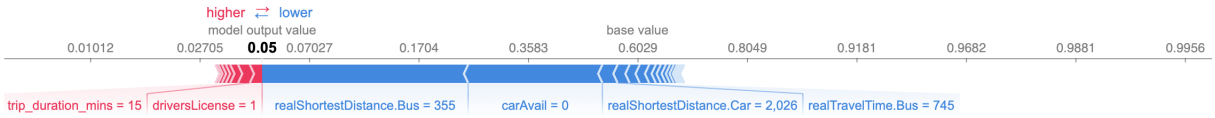

Figure 14: Positive and negative SHAP contributions (transformed to logit probability) for chosen female taking car travel mode, base value is decision boundary for class positive prediction [14]

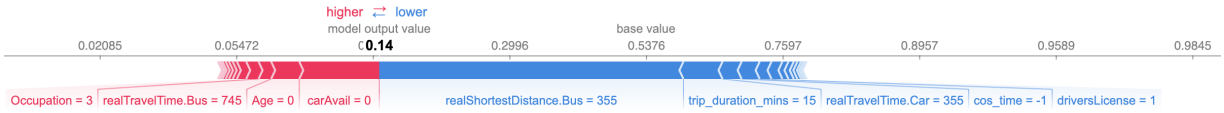

Figure 15: Positive and negative SHAP contributions (transformed to logit probability) for chosen female taking bus travel mode [15]

Figure 16: Positive and negative SHAP contributions (transformed to logit probability) for chosen female taking walk travel mode [16]

We observe from [16] that the model rightly classified this individual as walking with high confidence. [14] shows that whilst having a drivers licence is pushing the prediction toward the decision boundary, large negative SHAP values for distance and car availability push the output down. Such an analysis could be of great value to policy makers. Assuming this female is representative of other similar individuals in a student area, this could be used to understand how to optimally place infrastructure such as bike stations or bus stops, or which social norms

to target to influence behavioral change. Further, we can take this individual and randomly generate synthetic data for a feature of interest, to see for which threshold this feature may tip behaviour. [17][18] show where this threshold may be for two such features for this individual, in terms of how changes in these feature would contribute toward choosing to take a bus rather than walking. A small increase in real walking time would result in a small contribution toward choosing a bus. A 20 minute increase in trip duration would result in a significantly larger contribution to this individual choosing choosing a bus over walking.

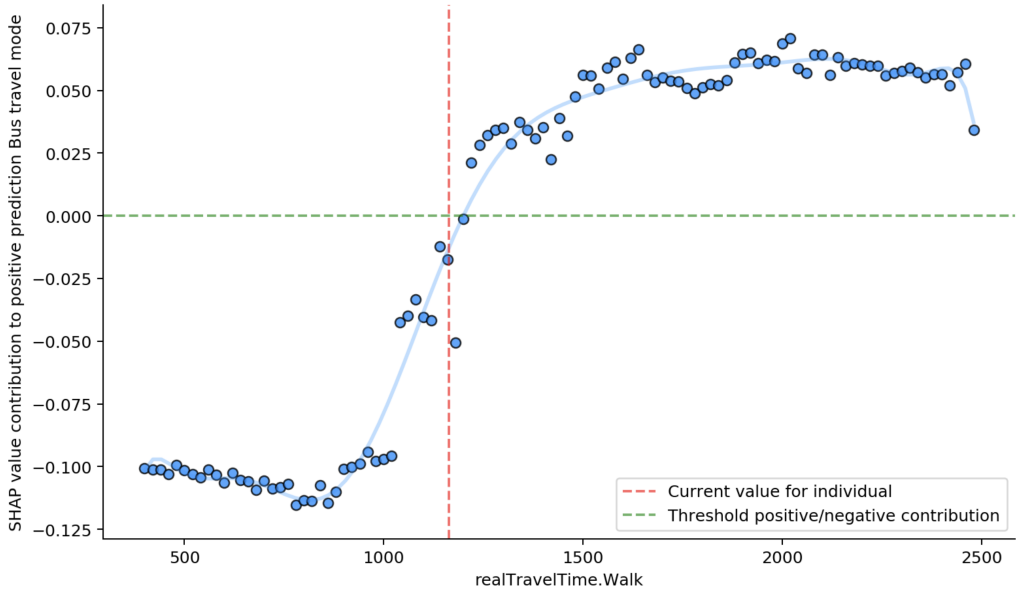

Figure 17: SHAP contribution from synthetically generated data for realTravelTime.Walk feature for individual in *table 4* [17]

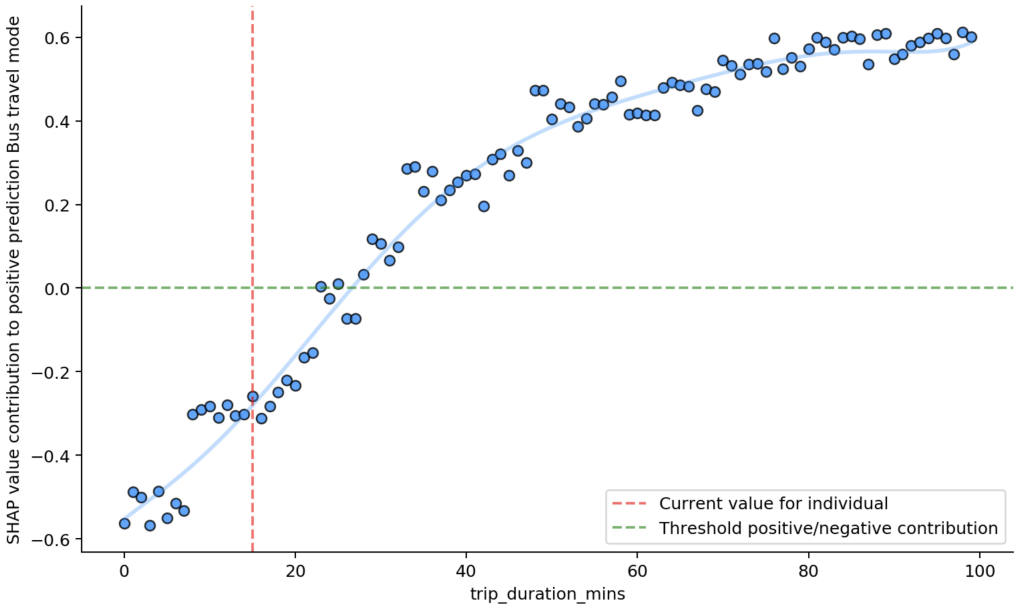

Figure 18: SHAP contribution from synthetically generated data for trip.duration.mins feature for individual in *table 4* [18]

VI Conclusion, limitations and future work

This paper demonstrates the applicability of using machine learning in mode choice analysis, and shows that a high degree of model interoperability can be derived, leading to meaningful insights which could help understand individual behaviour in the context of cities and transport. Whilst some of these insights could be derived from a traditional MNL model, this paper indicates that deeper machine learning models can achieve higher predictive accuracy, and hence inference closer to ground truth can be made. The study only provides a light touch treatment on the topic, deeper analysis should be made in the future, where more feature effects are explored. This study does not present the limitations to SHAP analysis or GBM models, for which there a number, this could receive greater treatment in further study. It would be useful to perform this analysis on a larger, more balanced dataset. Spatial correlations are not considered in this study. A natural evolution would be to incorporate the trip spatial constrains into trained models, either through mathematical graphs, or transport models.

References

Breiman, Leo (Aug. 2001). "Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)". In: *Statist. Sci.* 16.3, pp. 199–231. DOI: 10.1214/ss/1009213726. URL: https://doi.org/10.1214/ss/1009213726.

Chang, Ximing (2019). "Travel mode choice: a data fusion model using machine learning methods and evidence from travel diary survey data". In: *Transportmetrica A: Transport Science* 15.2, pp. 1587–1612. DOI: 10.1080/23249935.2019.1620380. eprint: https://doi.org/10.1080/23249935.2019.1620380. URL: https://doi.org/10.1080/23249935.2019.1620380.

Chen, Xianlong, Xiaoqian Liu, and Fazhi Li (2013). "Comparative study on mode split discrete choice models". In: *Journal of Modern Transportation* 21.4, pp. 266–272. DOI: 10.1007/s40534-013-0028-5. URL: https://doi.org/10.1007/s40534-013-0028-5.

Dubin, Jeffrey A. and Daniel L. McFadden (1984). "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption". In: *The Econometric Society*, pp. 345–362. URL: http://www.jstor.org/stable/1911493.

Gaker, D (2011). "Insights on Car-Use Behaviours from Behavioural Economics". In: *Emerald Group*.

London, Breathe (2020). *Breathe London - Map*. URL: https://www.breathelondon.org/. (accessed: 16.05.2020).

Lundberg, Scott and Su-In Lee (2017). *A Unified Approach to Interpreting Model Predictions*. arXiv: 1705.07874 [cs.AI].

Mcfadden, Daniel (1974). "Conditional logit analysis of qualitative choice behavior". In: URL: https://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf.

Molnar, Christoph (2019). *Interpretable Machine Learning - A Guide for Making Black Box Models Explainable*. URL: https://christophm.github.io/interpretable-ml-book/. (accessed: 26.05.2020).

Parr, Terence (Mar. 2016). *Beware Default Random Forest Importances*. URL: https://explained.ai/rf-importance/. (accessed: 26.05.2020).

Physicians, Royal College of (2019). *Every breath we take: the lifelong impact of air pollution*. URL: https://www.rcplondon.ac.uk/projects/outputs/every-breath-we-take-lifelong-impact-air-pollution. (accessed: 01.05.2020).

Sheffi, Yosef, Randy Hall, and Carlos Daganzo (1982). "On the estimation of the multinomial probit model". In: *Transportation Research Part A: General* 16.5, pp. 447–456. ISSN: 0191-2607. DOI: https://doi.org/10.1016/0191-2607(82)90071-1. URL: http://www.sciencedirect.com/science/article/pii/0191260782900711.

Wang, Fangru and Catherine L. Ross (2018). "Machine Learning Travel Mode Choices: Comparing the Performance of an Extreme Gradient Boosting Model with a Multinomial Logit Model". In: *Transportation Research Record* 2672.47, pp. 35–45. DOI: 10.1177/0361198118773556. eprint: https://doi.org/10.1177/0361198118773556. URL: https://doi.org/10.1177/0361198118773556.

Wang, Shenhao (2019). "PREDICTING TRAVEL MODE CHOICE WITH 86 MACHINE LEARNING CLASSIFIERS: AN EMPIRICAL BENCHMARK STUDY". In: URL: http://www.mit.edu/~baichuan/Baichuan/publication/Predicting.