EVALUATING THE PERFORMANCE AND INTEGRATION OF MACHINE LEARNING AND DISCRETE CHOICE MODELS ACROSS VARIOUS ASPECTS OF TRANSPORTATION PLANNING

Afia Ibnat Subah*[1] and Tanjila Rashid Rhythy[2]

[1] *Student, Bangladesh University of Engineering and Technology, Bangladesh,*
e-mail: aisubah181199@gmail.com
[2] *Student, Bangladesh University of Engineering and Technology, Bangladesh,*
e-mail: tanjila130@gmail.com

***Corresponding Author**

ABSTRACT

For several decades, discrete choice models (DCM) have been widely employed as a strategy for modeling travel mode choice. Various discrete choice models have been incorporated in many studies for predicting travel mode choices, such as Multinomial logit (MNL) model, Nested logit (NL)model, Heteroscedastic Extreme Value (HEV) Model, and Mixed logit model. However, the established conventional choice models are becoming outdated, raising concerns about their predictive precision. In contrast, machine learning (ML) techniques have emerged as a prominent approach in the transportation domain, exhibiting superior predictive capabilities compared to logit models. Machine Learning methods are gaining popularity with each passing day, permeating nearly every sphere of research, including the field of transportation engineering. Despite Machine Learning techniques not yet reaching full maturity for advanced predictions and occasionally exhibiting limitations in long-term forecasting due to inherent data-driven tendencies like overfitting, substantial progress is being achieved. Therefore, Machine Learning is surpassing traditional choice models in many instances. This paper presents a methodical assessment of the prediction performance of DCM and ML techniques across diverse scenarios of transportation planning that involves a comparative analysis between conventional choice models, specifically Multinomial logit models (MNL), and several prominent Machine Learning techniques such as Random Forests, Decision Trees, and Gradient Boosting trees, among others. Additionally, this paper highlights the increasing use of hybrid models which integrate machine learning and discrete choice models to effectively leverage the behavioural principles of traditional choice modelling with the predictive abilities of machine learning, suggesting their potential to improve overall modelling effectiveness.

Keywords: *Choice Model, Machine Learning, Overfitting, Random Forest, Gradient Boosting Trees*

1. INTRODUCTION

Transportation planning is a process which defines the policies, goals and investments that are needed to move people and goods efficiently, economically and in the fastest way possible which should be an inclusive and holistic process involving all communities and groups impacted by transportation infrastructure (Palamariu & Tulbure, 2021). A *paradigm shift* from mobility-based analysis to accessibility-based analysis is occurring in transportation planning (Litman, 2008). As a result, it is critical to efficiently model transportation planning in the context of current trends, for which various methods, including discrete choice modeling and machine learning techniques, are widely used.

The discrete choice model, a type of disaggregate model with advantages over aggregated models, can be used to analyze and predict a decision maker's selection of one alternative from a finite set of mutually exclusive and collectively exhaustive alternatives (Koppelman & Bhat, 2006). Discrete choice models have widespread use in transportation since McFadden (1974) developed random utility framework. It has been applied in travel mode choice, destination mode choice (Bhat et al., 1998), route choice (Ben-Akiva et al., 2004; Yai et al., 1997), air travel choices (Proussaloglou & Koppelman, 1999), activity analysis (Wen & Koppelman, 1999). Commonly used discrete choice models include Multinomial Logit Model (MNL), Nested Logit Model (NL), Heteroscedastic Extreme Value (HEV)Model, Mixed logit model etc.

Because of significant advances in machine learning methods, machine learning has become a popular method in the transportation sector. In ML terminology, Random Utility Model can be considered as a supervised probabilistic classifier; the aim of the model is to predict the probability of an individual choosing each mode (i.e., the classes), given a set of features (variables) describing the choice situation. The modeller has access to a finite dataset of choice situations alongside the ground-truth class labels (the option chosen) to train the model (Hillel et al., 2021). The common machine learning techniques used for choice modelling include Logistic Regression (LR), Artificial Neural Networks (ANNs), Decision Trees (DTs), Ensemble Learning (EL), and Support Vector Machines (SVMs).

Despite the growing popularity of machine learning techniques, many researchers still prefer traditional discrete choice models because machine learning techniques rely heavily on data. The purpose of this paper is to,

- Compare two widely used techniques in the contexts of travel behavior analysis, vehicle ownership selection, and freight vehicle type selection.
- Discuss which model is superior to the other one.
- Give an overview of the models that incorporate both approaches.

We hope to demonstrate in this paper that, because both models have flaws, it is best to combine them in order to fully reap the benefits of both the theory-driven and data-driven approaches.

2. TRAVEL BEHAVIOR ANALYSIS

Although random utility modelling has been popular in determining the parametric relationships between mode choice and its possible determinants since the 1980s, there are some underlying issues with this method. This method makes the assumption that each choice is independently and identically distributed (IID), which can lead to inaccurate predictions. Flexible random utility model methods, such as the mixed logit model, have the advantage of having relaxed IID assumptions, which results in better modeling performance. Nonetheless, it has predetermined structures and linear properties of underlying functions, making it difficult to capture a high degree of non-linearity (Lee et al., 2018).

In contrast to logit models, which capture non-linearity by reducing the complexity of the dataset, ANNs are exceptionally effective in capturing non-linear properties with the help of additional units or hidden layers. Nonetheless, ANNs have their own set of drawbacks, such as a lack of interpretability and a relative inability to use previously acquired knowledge. Despite these disadvantages, ANNs outperform logit models (Lee et al., 2018).

Lee et al. (2018) compare four types of artificial neural networks (ANNs) and the Multinomial Logit Model (MNL) in the context of travel mode choice modelling using the Revealed Preference (RP) survey of the Chicago Metropolitan Area for Planning (CMAP) Travel Tracker Survey dataset collected from 2007 to 2008. Back Propagation Neural Network (BPNN), Radial Basis Function Neural Network (RBFN), Probabilistic Neural Network (PNN), and Clustered Probabilistic Neural Network (CPNN) are the four types of ANN models. According to the results, all four ANNs have better overall model accuracy, which is around 80%, than the MNL, which has a model accuracy of 70.5 percent.

Sehmisch (2021) examined the utilization of Artificial Intelligence, particularly Machine Learning classifiers, in comparison to classical Logit models for trip decision modeling. Although Machine Learning exhibits robust prediction capabilities, it frequently lacks interpretability. The study compared Logit models with two AI algorithms, namely Artificial Neural Networks and Decision Trees, focusing on evaluating the predictive accuracy of these models and their capacity to offer significant economic insights. The results indicated that Neural Networks provide credible metrics, albeit distinct from Logit models, whereas the Classification Tree algorithm is less appropriate. Machine Learning methods surpass Logit models in prediction performance, indicating a tradeoff in methodology based on modeling aims.

Salas et al. (2022) conducted a comparison of traditional discrete choice models, specifically multinomial logit (MNL) and mixed multinomial logit (MMNL), and five machine learning classifiers in order to represent travel mode choice. In addition to evaluating predicted accuracy, this study also evaluated the ability to explain results after the fact using synthetic datasets. The results indicated a decrease in the disparity of accuracy when taking into account variations in taste preferences. Neural Networks surpass other models in terms of both accuracy and interpretation, highlighting the importance of conducting model equivalence analysis to improve decision-making assistance and comprehend the elements influencing trip choices.

Rahnasto (2022) explored a comparative analysis between machine learning and classical discrete choice models to forecast travel destination preferences. When evaluating models for different sorts of trip activities and modes, the random forest model performs better than the others in overall ratings. Transitioning from multinomial logit to random forest enhances prediction accuracy by up to 40% in binary evaluations. When dealing with larger amounts of data, gradient boosted regression, binomial logit, and neural network models outperform the multinomial logit model. These findings have important consequences for predicting travel demand and choosing destinations. They emphasize the potential of specific machine learning methods to improve the accuracy of predictions in cases when destination choices are uncertain.

Wang et al. (2021) conducted a thorough evaluation of machine learning (ML) classifiers and discrete choice models (DCMs) in order to forecast travel behavior. The results showed that ensemble approaches and deep neural networks have superior predictive capabilities, while random forests strike a favorable trade-off between prediction accuracy and computational efficiency. DCMs exhibit competitive accuracy but are prone to lengthy computational durations, particularly when dealing with big samples or high dimensionality. The comparative order of classifiers remains consistent, while the exact values differ. This research suggested utilizing deep neural networks, model ensembles, and random forests as fundamental models for predicting future travel behavior.

Püschel et al. (2023) examined the interconnectedness of different mobility tools in individual travel behavior, namely mode choice, which is a factor that is not taken into account in conventional sequential logit models. The research demonstrated that when applied to a synthetic population, both discrete choice and neural network models more accurately approach target distributions than sequential logit models. The shallow and deep neural networks that were evaluated exhibit superior prediction accuracy, with the networks including only one hidden layer being more resilient and simpler to conceive and comprehend compared to the deeper networks.

Zhao et al. (2019) undertook a thorough comparison of logit and machine-learning models, analyzing the similarities and differences in the construction, evaluation, and interpretation of behavioral patterns for travel mode choice modeling. An empirical assessment of a stated-preference survey dataset demonstrated that machine learning surpasses logit models in terms of predicting accuracy. However, both approaches generally concur on behavioral interpretations.

Through a thorough comparison of the behavioral outputs and forecast accuracy of machine learning —more specifically, random forest—with conventional logit models, (Zhao et al., 2020) filled a gap in the travel behavior study. The random forest model had been empirically evaluated using stated-preference survey data, which validated its better prediction accuracy. The research highlighted initial difficulties encountered with the random forest model, which produced behaviorally implausible outcomes. However, these obstacles were successfully addressed by making improvements that specifically targeted the constraints of tree-based models.

Kamkar (2021) explored the commuting patterns of individuals affiliated with the University of Calgary, discovering that 57.89% utilize environmentally friendly means of transportation. Socio-demographic characteristics exert a substantial influence on travel behavior, with age and income playing a crucial role in determining the use of cars and public transit. The study offered valuable insights about satisfaction levels, obstacles, and concerns associated with various forms of transportation, and presents policy recommendations to promote sustainable mobility. Extreme Gradient Boosting (XGBoost), a type of machine learning classifier, demonstrated superior performance compared to regular multinomial logit models when it comes to predicting transportation modes. The study determined that travel-related data has a greater influence on machine learning methods, whereas socio-demographic variables play a critical role in multinomial logit models.

N. F. M. Ali et al. (2021) aimed to evaluate and compare the performance of various machine learning models (Neural Network, Random Forest, Decision Tree, and Support Vector Machine) with the conventional Discrete Choice Model (Binary Logistic Regression) in predicting travel mode selection in Kuantan City, Malaysia. The Neural Network model exhibited superior performance compared to Binary Logistic Regression when utilizing Revealed/Stated Preferences (RP/SP) Survey data, obtaining a higher prediction accuracy of up to 73.4% for training and 72.4% for testing. Feature importance analysis found the crucial aspects that have a major impact on the decision of choosing a travel mode. The study highlighted the capacity of machine learning models, namely Neural Networks, to improve urban transportation planning. It also recognized the significance of the Discrete Choice Model in comprehending the connections between variables to enhance future transportation systems.

3. VEHICLE OWNERSHIP CHOICES

Ali et al. (2023) compared Machine Learning (ML) such as Neural Network (NN), Gradient Boosting Trees (GBT), and Choice Model (CM) in the context of vehicle ownership using household survey data from developing countries from three different years. They created vehicle ownership models for both backcasting and forecasting. The models are compared based on their Log Likelihood and Mean Absolute Percentage Error (MAPE). 20% of the data was used for testing, while the other 80% was used for training and cross-validation (in the case of ML). With a higher value of LL, the MNL outperforms both Gradient Boosting trees and Neural Networks in terms of prediction performance. MNL has the lowest MAPE value yet again. When it comes to weighted MAPE, MNL has the lowest value yet again. However, whereas the different income groups were directly incorporated into the utility function of the MNL model, the ML model learns it "automatically," which reduces the likelihood of model misspecification in ML models because it does not rely on the modeller's intuition and specification. As a result, ML techniques should be used for forecasting because they can capture different market segments. Furthermore, while MNL models outperformed neural networks in forecasting even when the temporal gap was large and the landscape had changed significantly, neural networks outperformed MNL models in back casting when the temporal gap was short, and the landscape had not changed significantly. As a result, the study concludes that it is difficult to determine which method performs better in general and in all contexts.

Using transportation household survey data from Singapore, (Paredes et al., 2017) compared household car ownership models (Multinomial Logit Model) with different machine learning models such as Random Forest, Support Vector Machines - SVMs, Decision trees, Extreme Gradient Boosting, and an Ensemble of methods. They used 2008 data to train the ML models and estimate the MNL models, which they then used to predict car ownership in 2012. According to the findings, while ML models underperformed on prediction when using the dataset prepared for the MNL model, they outperformed the MNL model by 10% after applying some feature engineering. As a result, both types of models necessitate some form of data pre-processing in order to be fully utilized.

4. CHOICE OF FREIGHT VEHICLE TYPE

Ahmed & Roorda (2022) compared the random forest model to the multinomial and mixed logit models in the context of freight vehicle type because vehicle type selection is one of the most important logistics decisions that firms make. They modeled the selection of four road transport vehicle types because, in an urban area, road transport is the only available mode of transportation (Jong, 2013). The most common method for modeling freight mode choice is the discrete choice model. However, Ahmed & Roorda (2022) stated that their study is the first to use machine learning algorithms to model freight vehicle type selection. For the development of the models, they used a commercial travel survey with information about outbound shipment transportation in the Toronto area.

In terms of prediction accuracy, the study's findings show that the Random Forest model outperforms both the MNL and mixed MNL models. As a result, the study suggests that the RF model be "further developed and potentially applied" (Ahmed & Roorda, 2022).

5. WHICH APPROACH IS SUPERIOR?

Although Machine Learning models are relatively new, they have enormous potential for use in transportation engineering. Because it provides a data-driven approach, as opposed to traditional Discrete Choice Models, which are primarily theory-driven, it is more appropriate to use and accept Machine Learning models now than ever in the era of readily available data. However, there are some common misconceptions about Machine Learning, such as the fact that machine learning models can only be used for prediction rather than behavioral inference, and that machine learning models frequently overfit the data. Third, as a result of the preceding two factors, there appears to be a lack of recognition of the potential value of integrating machine learning models, techniques, and practices for the choice modeling field (Van Cranenburgh et al., 2022).

It is critical to understand that both models have limitations that impede their performance. As a result, determining which approach is superior to the other is difficult. Nonetheless, because the primary goal of a modeler should be to maximize model performance, crossover of both models appears to be a very good idea in recent years, as (Van Cranenburgh et al., 2022) have showed that Machine Learning Models can be very effective in overcoming the problems and limitations of the traditional Discrete Choice approach, such as subjective labor-intensive search processes for model selection, and the inability to work with text and image data.

6. COMBINATION OF MACHINE LEARNING AND DISCRETE CHOICE MODELS

While A. Ali et al. (2023) primarily focused on comparing traditional discrete choice models (CM), specifically multinomial logit (MNL), with machine learning techniques (ML) in the context of vehicle ownership choices in Dhaka, Bangladesh, this research also hinted at the growing interest in combining these approaches. Recent developments like Learning Multinomial Logit, TasteNet-MNL, Embeddings Multinomial Logit, and ASU-DNN model, which are hybrids of MNL and neural networks (NN), incorporating both traditional choice modeling and machine learning.

This study suggested that future research could further explore and compare the performance of these hybrid models with conventional MLs and CMs. This hybridization may offer a potential avenue for leveraging the strengths of both approaches, combining the behavioral underpinning of traditional models with the predictive power of machine learning, thereby enhancing the overall effectiveness of transportation modeling.

Arkoudi et al. (2023) presented a new method that integrates theory- and data-based choice models using Artificial Neural Networks (ANNs), with a specific emphasis on the capacity to understand and interpret the model. This approach differs from prior research by emphasizing the interpretability of the embedding vectors. It achieves this by linking each dimension of the vectors to a specific option alternative, resulting in outputs that have clear behavioral significance. The primary model, Embeddings Multinomial Logit (E-MNL), which is built on artificial neural network (ANN) principles, maintained the capacity to interpret utility coefficients for all input variables while achieving exceptional predictive accuracy. When tested on actual datasets, the suggested models performed better than existing artificial neural network (ANN) models. They showed enhanced efficiency by employing fewer network parameters compared to baseline models that use dummy encoding.

7. CONCLUSION AND FUTURE RECOMMENDATIONS

In conclusion, this paper highlights the evolving characteristics of transportation modelling, contrasting the traditional discrete choice models (DCM) with the increasing importance of machine learning (ML) techniques. The evaluation of different methodologies, namely in the analysis of travel behaviour, decisions regarding vehicle ownership, and selection of freight vehicle types, exposes the subtle advantages and constraints of each technique.

This paper indicates that although ML models frequently demonstrate higher prediction abilities, they are not devoid of obstacles, such as interpretability and reliance on data. Therefore, the question of determining which methodology is best remains intricate, resulting in an increasing interest in integrating both methodologies.

Emerging as potential answers are recent hybrid models such as Learning Multinomial Logit, TasteNet-MNL, Embeddings Multinomial Logit, and ASU-DNN. These models aim to combine the behavioural foundations of traditional models with the predictive capabilities of machine learning. The integration, as emphasised in the literature, has the potential to improve the overall efficiency of transportation modelling by tackling the constraints of different methodologies.

Following future recommendations can be made from this study:

- Further research should focus on exploring the performance and applicability of hybrid models that combine traditional discrete choice models (DCM) with machine learning (ML) techniques. Some examples of these hybrid models include Learning Multinomial Logit, TasteNet-MNL, Embeddings Multinomial Logit, and ASU-DNN. Studying hybrid approaches in different transportation scenarios can offer valuable insights into their effectiveness and limitations.

- The interpretability of machine learning (ML) models, particularly neural networks, has been subject to criticism. To address this concern, future research should prioritize the development of ML models that achieve a balance between predictive accuracy and interpretability. This may entail developing new methods or adjusting current models to improve their capacity in offering valuable insights for decision-making.

- Addressing the limitations of machine learning techniques in long-term forecasting, such as overfitting and data-driven tendencies, is crucial. Future research could aim to enhance ML algorithms for more accurate prediction of transportation choices over longer time periods, ensuring their reliability for future planning scenarios.

- Conducting comparative studies across diverse contexts, such as different geographical locations, cultural settings, and demographic compositions, can enhance our understanding of the generalizability and reliability of discrete choice models and machine learning techniques. Analyzing the performance of these models across various regions and populations can aid in identifying influential factors.

- The integration of qualitative and quantitative data is necessary in transportation decision-making as quantitative models alone may not fully capture the influence of qualitative factors on transportation choices. Future research should investigate methods for incorporating qualitative data into the modeling process to improve the accuracy and comprehensiveness of predictions. This may entail integrating survey data with subjective insights obtained from interviews or focus groups.

- Future research could prioritize the development of dynamic models that can adapt to changing conditions in transportation systems. This involves studying ways to update transportation models in real-time or near-real-time, to ensure their relevance and accuracy in the face of dynamic societal, economic, and technological changes.

- Transportation choices are influenced by external factors such as policy changes, economic shifts, and technological advancements. Future research should investigate methods for integrating external factors into models to improve their predictive abilities and facilitate better decision-making in transportation planning.

- Promoting open data standards and fostering collaboration among researchers and practitioners can facilitate the sharing of datasets, methodologies, and findings. This collaborative approach can enhance the development of transportation models, promoting progress in the field.

To summarize, the comparison and combination of discrete choice models and machine learning techniques in transportation modeling indicates a promising direction for future research. This highlights the importance of adopting a balanced approach that takes into account both predictive accuracy and interpretability to tackle the challenges and opportunities in this evolving field.

8. REFERENCES

Ali, A., Kalatian, A., & Choudhury, C. F. (2023). Comparing and contrasting choice model and machine learning techniques in the context of vehicle ownership decisions. Transportation Research Part A: Policy and Practice, 173, 103727. https://doi.org/10.1016/j.tra.2023.103727

Ali, N. F. M., Sadullah, A. F. M., Majeed, A. P. A., Razman, M. A. M., Zakaria, M. A., & Nasir, A. F. A. (2021). Travel Mode Choice Modeling: Predictive Efficacy between Machine Learning Models and Discrete Choice Model. The Open Transportation Journal, 15(1). https://doi.org/10.2174/1874447802115010241

Arkoudi, I., Krueger, R., Azevedo, C. L., & Pereira, F. C. (2023). Combining discrete choice models and neural networks through embeddings: Formulation, interpretability and performance. Transportation Research Part B: Methodological, 175, 102783. https://doi.org/10.1016/j.trb.2023.102783

Ben-Akiva, M. E., Ramming, M. S., & Bekhor, S. (2004). Route Choice Models. In M. Schreckenberg & R. Selten (Eds.), Human Behaviour and Traffic Networks (pp. 23–45). Springer Berlin Heidelberg.

Bhat, C., Govindarajan, A., & Pulugurta, V. (1998). Disaggregate Attraction-End Choice Modeling: Formulation and Empirical Analysis. Transportation Research Record, 1645(1), 60–68. https://doi.org/10.3141/1645-08

Hillel, T., Bierlaire, M., Elshafie, M. Z. E. B., & Jin, Y. (2021). A systematic review of machine learning classification methodologies for modelling passenger mode choice. Journal of Choice Modelling, 38, 100221. https://doi.org/10.1016/j.jocm.2020.100221

Kamkar, H. (2021). Travel Behavior Analysis and Mode Choice Prediction for Commuting to Campus—Performance Comparison of Discrete Choice and Machine Learning Models. http://hdl.handle.net/1880/113922

Koppelman, F., & Bhat, C. (2006). A Self Instructing Course in Mode Choice Modelling: Multinomial and Nested Logit Models.

Lee, D., Derrible, S., & Pereira, F. C. (2018). Comparison of Four Types of Artificial Neural Network and a Multinomial Logit Model for Travel Mode Choice Modeling. Transportation Research Record: Journal of the Transportation Research Board, 2672(49), 101–112. https://doi.org/10.1177/0361198118796971

Litman, T. (2008). Evaluating accessibility for transportation planning. Victoria Transport Policy Institute, Victoria, Canada.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), Fontiers in Econometrics (pp. 105–142). Academic press.

Palamariu, M., & Tulbure, I. (2021). Chapter 3—Land use and management strategies for shaping sustainable cities. In M.-M. Nistor (Ed.), Climate and Land Use Impacts on Natural and Artificial Systems (pp. 29–48). Elsevier. https://doi.org/10.1016/B978-0-12-822184-6.00004-1

Proussaloglou, K., & Koppelman, F. S. (1999). The choice of air carrier, flight, and fare class. Journal of Air Transport Management, 5(4), 193–201. https://doi.org/10.1016/S0969-6997(99)00013-7

Püschel, J., Barthelmes, L., Kagerbauer, M., & Vortisch, P. (2023). Comparison of Discrete Choice and Machine Learning Models for Simultaneous Modeling of Mobility Tool Ownership in Agent-Based Travel Demand Models. Transportation Research Record, 03611981231206175. https://doi.org/10.1177/03611981231206175

Rahnasto, I. (2022). Comparing discrete choice and machine learning models in predicting destination choice. https://aaltodoc.aalto.fi/handle/123456789/115174

Salas, P., De la Fuente, R., Astroza, S., & Carrasco, J. A. (2022). A systematic comparative evaluation of machine learning classifiers and discrete choice models for travel mode choice in the presence of response heterogeneity. Expert Systems with Applications, 193, 116253. https://doi.org/10.1016/j.eswa.2021.116253

Sehmisch, S. (2021). Comparative Choice Analysis using Artificial Intelligence and Discrete Choice Models in A Transport Context.

Van Cranenburgh, S., Wang, S., Vij, A., Pereira, F., & Walker, J. (2022). Choice modelling in the age of machine learning—Discussion paper. Journal of Choice Modelling, 42, 100340. https://doi.org/10.1016/j.jocm.2021.100340

Wang, S., Mo, B., Hess, S., & Zhao, J. (2021). Comparing hundreds of machine learning classifiers and discrete choice models in predicting travel behavior: An empirical benchmark (arXiv:2102.01130). arXiv. https://doi.org/10.48550/arXiv.2102.01130

Wen, C.-H., & Koppelman, F. S. (1999). Integrated Model System of Stop Generation and Tour Formation for Analysis of Activity and Travel Patterns. Transportation Research Record, 1676(1), 136–144. https://doi.org/10.3141/1676-17

Yai, T., Iwakura, S., & Morichi, S. (1997). Multinomial probit with structured covariance for route choice behavior. Transportation Research Part B: Methodological, 31(3), 195–207. https://doi.org/10.1016/S0191-2615(96)00025-2

Zhao, X., Yan, X., Yu, A., & Van Hentenryck, P. (2019). Modeling Stated Preference for Mobility-on-Demand Transit: A Comparison of Machine Learning and Logit Models (arXiv:1811.01315). arXiv. https://doi.org/10.48550/arXiv.1811.01315

Zhao, X., Yan, X., Yu, A., & Van Hentenryck, P. (2020). Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. Travel Behaviour and Society, 20, 22–35. https://doi.org/10.1016/j.tbs.2020.02.003