

Euro Working Group on Transportation Annual Meeting 2025 - EWGT2025

Understanding Commuter Mode Choice in Rome: A Comparative Analysis of Neural Networks and Multinomial Logit for Sustainable Mobility

Marzieh Afsari^{1*}, Ken Koshy Varghese¹, Lory Michelle Bresciani Miristice¹, Guido Gentile¹

¹*Sapienza University, via Eudossiana, 18 - 00184 Rome, Italy*

Abstract

Understanding how commuters select their travel modes is vital for advancing sustainable urban mobility and reducing reliance on private vehicles. This study compares the performance of Multinomial Logit (MNL) and Neural Network (NN) models using travel survey data collected from employees in Rome, Italy. Both models are evaluated under a consistent preprocessing and validation pipeline to ensure fair comparison. While the NN model achieves slightly higher predictive accuracy, a paired t-test confirms that the performance difference is not statistically significant. Elasticity analysis and SHAP-based interpretation are employed to assess the impact of key variables. SHAP results show that both models highlight factors such as public transport inefficiency, cost, and parking difficulty. The MNL model offers sharper prioritization of key features, while the NN reveals broader, non-linear patterns. These findings support hybrid approaches that combine MNL interpretability with NN flexibility to improve behavioral insights and guide sustainable transportation policies.

© 2026 The Authors. Published by ELSEVIER B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the Euro Working Group on Transportation Annual Meeting 2025 - EWGT2025.

Keywords: SHAP value; Elasticity analysis; Model interpretability; Policy insights.

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .

E-mail address: Marzieh.afsari@uniroma1.it

1. Introduction

Commuters play a critical role in shaping traffic flow, representing a substantial share of daily travellers. In contrast to non-commuters, who have more flexible travel behaviour, commuters follow consistent long-term travel patterns dictated by work or school obligations. Understanding commuting patterns is essential for effective transportation planning and sustainable policy development. While some research focuses on optimizing traffic flow and congestion management through network-level models, a more sustainable approach involves encouraging behavioural shifts toward active and public transportation (Tabatabaei et al., 2023). Policies that promote alternative modes, rather than solely improving traffic efficiency, can lead to long-term environmental and social benefits.

Traffic congestion not only undermines transportation efficiency but also affects individual well-being, contributing to increased stress and reduced job satisfaction. In this context, understanding and influencing mode choice becomes essential. Mode choice modelling serves as a key tool for analysing and predicting travel behaviour, enabling researchers and policymakers to evaluate how various factors, such as travel time, cost, and personal preferences, influence the selection of transport modes. Traditional Discrete Choice Models (DCMs) are widely appreciated for their strong theoretical foundation and ability to provide interpretable insights into individual decision-making processes (Hillel et al., 2021). In Machine Learning (ML) terms, a Random Utility Model (RUM) can be interpreted as a supervised probabilistic classifier, predicting mode choice probabilities from a finite dataset with ground-truth labels (Afsari et al., 2025). ML classification algorithms, known for their effectiveness, are widely applied in transportation tasks such as safety assessment and demand prediction (Eldafrawi et al., 2024; M. Afsari et al., 2024), capture complex, non-linear patterns without requiring predefined utility specifications, offering greater flexibility than RUMs. While ML algorithms often achieve higher predictive accuracy, DCMs remain valuable for their transparency and policy relevance.

To bridge the gap between prediction and interpretation, Shapley values (Shapley, 1953) have emerged as a powerful method to assess the contribution of explanatory variables in both classical and machine learning (ML) models. Originating from cooperative game theory, the Shapley value provides a fair allocation of a total outcome (such as model accuracy or fit) across all features by averaging their marginal contributions over all possible subsets. This method has been adapted to different contexts, including discrete choice models and generalized linear models.

(Lundberg & Lee, 2017) extended this concept to ML with SHAP (SHapley Additive exPlanations), a practical and efficient method that provides both local (individual prediction-level) and global (model-level) interpretability. SHAP is now widely used in transportation and mobility research to interpret complex model outputs. Its solid theoretical foundation and consistent feature attributions make it suitable for explaining the relative importance of variables in both traditional and ML-based mode choice models.

Given the complementary strengths of classical and ML models as well as the increasing use of interpretability tools like SHAP, selecting the appropriate model depends heavily on the dataset characteristics and the research objectives (García-García et al., 2022). Since no single model is universally superior, this study aims to compare Multinomial Logit (MNL) and Neural Networks (NN) to analyse their differences in prediction accuracy, feature importance, and policy implications. While MNL provides interpretability, offering clear insights into the factors influencing mode choice, NN excels at capturing complex, non-linear relationships. By evaluating both models, we aim to determine their relative effectiveness in predicting commuting behaviour and informing sustainable transportation policies in Rome, Italy. This study makes several key contributions:

- *Comparative Analysis of ML and Econometric Models:* Systematically compares MNL and NN models for mode choice prediction, addressing existing methodological gaps highlighted in prior literature; ensures consistency through a structured preprocessing pipeline, cross-validation, and benchmarking for rigorous model comparison.
- *Enhanced Interpretability:* SHAP analysis applies to both MNL and NN to enable a comparison of feature importance. Additionally, it examines elasticity to quantify the impact of key variables on commuting choices. By integrating these approaches, we bridge the gap between predictive accuracy and interpretability, ensuring a more comprehensive understanding of model insights for transportation policy.
- *Sustainable Mobility Insights:* Examines the key factors influencing the transition to sustainable transportation modes and provide evidence-based policy recommendations to encourage alternative mobility options and reduce car dependency.

2. Methodology

An MNL and an NN model were developed using a structured pipeline for fair comparison, as illustrated in Fig 1. Both models use the same preprocessing steps, train-test splitting, and evaluation metrics. They undergo confusion matrix evaluation and out-of-sample accuracy assessment, ensuring a consistent evaluation framework.

2.1. MNL Model

Unordered choice models are based on utility maximization theory, where individuals select the alternative that provides the highest utility. Since utility is not directly observable, the random utility theory is applied (Ben-Akiva & Lerman, 1985), decomposing utility into a deterministic component and a random error term. The deterministic component $V_{n,i}$ captures observable attributes of the alternative i and individual characteristics n and is expressed as $V_{n,i} = f(\mathbf{B}, \mathbf{x}_{n,i})$, where \mathbf{B} represents the estimated parameters and $\mathbf{x}_{n,i}$ the explanatory variables. When error terms follow an independently and identically distributed (IID) Type 1 Extreme Value distribution, the probability $P_{n,i}$ of individual n selecting alternative i is given by the MNL model:

$$P_{n,i} = \frac{e^{V_{n,i}}}{\sum_{j=1}^J e^{V_{n,j}}} \tag{1}$$

The \mathbf{B} parameters are determined by maximizing the following log-likelihood (LL) function, where $y_{(n,i)}$ is 1 if the decision-maker n chooses the alternative i :

$$LL(\mathbf{B}) = \sum_{n=1}^N \sum_i y_{n,i} \ln(P_{n,i}) \tag{2}$$

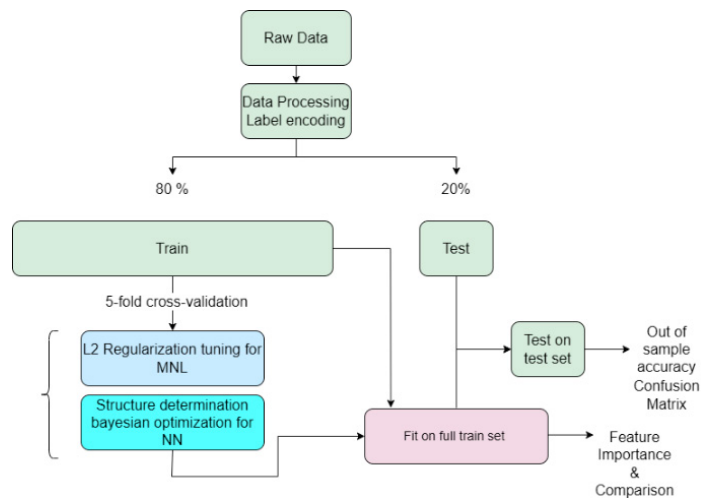


Fig 1: MNL and NN modelling workflows for mode choice prediction

In this study, mode choice probability is modelled as a function of 46 independent variables, assuming a linear utility function within the random utility maximization framework. The model applies 5-fold cross-validation to fine-tune L2 regularization. Model estimation is performed using the Newton algorithm with a trust region for simple bound

constraints to ensure efficient convergence. The analysis is conducted in Python using the Biogeme library (Bierlaire, 2025), a widely used tool for discrete choice modelling.

2.2. NN Model

NN operates as a system of linear equations where weights (\mathbf{W}) and biases (\mathbf{b}) connect neurons, with each neuron undergoing a nonlinear transformation as described in Equation (3). Here, Z_i represents the logit score of class i . The network parameters are optimized using the maximum likelihood principle, minimizing cross-entropy or maximizing LL. In multi-class classification, the final layer typically uses a softmax activation function (Equation (4)), which normalizes outputs between 0 and 1, ensuring that the predicted probabilities sum to 1 across all choices.

$$Z_i = f(W * x + b) \quad (3)$$

$$P(i | x) = \frac{e^{z_i}}{\sum_i e^{z_i}} \quad (4)$$

To identify the best-performing architecture, Bayesian search was employed to tune hyperparameters, including learning rate, number of hidden layers and their sizes, activation functions, regularization strength (L2 alpha), and optimizer settings. This method systematically balances exploration and exploitation to efficiently search the hyperparameter space. As a result, the final network architecture consists of an input layer followed by two hidden layers with 128 and 64 neurons, respectively, using the hyperbolic tangent (tanh) activation function. The output layer employs a softmax activation to classify commuting modes into five classes: car, motorcycle, walking, bicycle, and public transport (PT). The model is trained using stochastic gradient descent with an adaptive learning rate, and regularization ($\alpha = 0.1$) is applied to mitigate overfitting. The model's generalizability was validated using 5-fold CV.

2.3. Shapley Values

To enable a consistent comparison of feature importance across the MNL and NN models, we apply the SHAP (SHapley Additive exPlanations) framework, which assigns each feature i a value ϕ_i representing its average marginal contribution to the model output (Lundberg & Lee, 2017). The model prediction is approximated as:

$$f(x') = \phi_0 + \sum_{i=1}^M \phi_i x'_i \quad (5)$$

where ϕ_0 is the expected model output, and x'_i indicates whether feature i is present. While prior work (Salas et al., 2025) introduced Shapley-based analysis for MNL using McFadden's pseudo- R^2 (McFadden, 1973), we apply the SHAP library directly to both models to generate comparable global importance values. This allows us to assess how each model attributes explanatory power to features within a unified interpretability framework.

3. Results and Discussion

3.1. Data Preprocessing

The study uses travel mode choice survey data from 2023, collected across employees of 30 companies in the Rome metropolitan area. After data cleaning, the final dataset included 1,688 valid responses, with 44 variables available for analysis. The dataset includes five primary commuting modes (car, motorcycle, PT, walking, and bike), with travel times calculated using the Google Maps API based on employees' recorded departure times. These calculations account for real-time traffic congestion during morning peak hours. To ensure consistency, Wednesdays (20th Nov 2024) were chosen as the reference day for analysis. Flexible workers avoid peak congestion, so travel time was the main measure used to assess their commuting patterns. Motorcycle travel time, not in Google Maps, was estimated at 90% of car travel, reflecting a 10% based on observed urban speed differences (Buchanan & Walton, 2012).

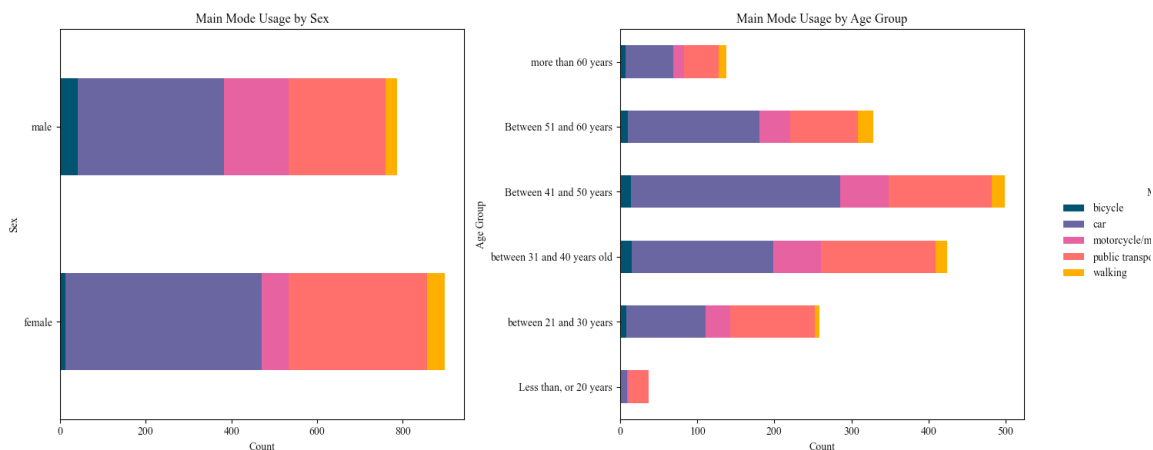


Fig 2: Mode choice distribution by (a) sex, (b) age groups.

The distribution of transportation mode usage varies by gender and age (Fig 2). Both men and women primarily use cars and motorcycles for commuting, with a notable portion relying on public transport, while walking and cycling remain minimal. Younger individuals, especially those under 30, are more likely to use PT, whereas those over 40 show a stronger preference for cars.

3.2. Comparing the MNL and NN Performance

The models include 44 features grouped into the following categories:

- **Socio-demographic attributes:** Age, gender, and family structure.
- **Work-related attributes:** Contract type, employment category, work mission, number of working days per month, and daily working hours.
- **Mobility-specific attributes:** Travel time and cost for cars, motorcycles, PT, and travel time for walking, and bike; number of PT transfers, walking time to/from PT stops, subscription cost, and access mode to PT stops. Parking-related variables (availability, fee, and difficulty) are also included.
- **Time-related flexibility:** Flexibility of entry time, lunch break activity, and presence of intermediate stops during the trip.
- **Social and behavioral aspects:** Whether the person travels with others, reasons for mode choice (e.g., cost, comfort, health, independence, environment), and alternative modes used in other seasons.

Table 1 presents the performance comparison between the MNL and the NN models using balanced accuracy, macro precision, macro recall, and macro F1-score. The NN model achieves slightly higher results and suggests better performance in capturing distinctions across all commuting modes. To assess whether this difference is statistically significant or due to randomness, a paired t-test was conducted on precision values from cross-validation folds. A paired t-test on precision scores obtained from the same cross-validation folds produced a two-tailed p-value of 0.108, indicating no statistically significant difference between the models. This suggests that for medium-sized datasets, a well-implemented MNL model remains a viable alternative despite the NN model’s slight advantage.

Table 1: Comparison of MNL and Neural Network models on classification performance metrics.

Method	Accuracy (Balanced)	Precision (Macro)	Recall (Macro)	F1 score (Macro)
MNL	0.6915	0.6771	0.6915	0.6748
NN	0.7172	0.7263	0.7172	0.7154

To better understand the influence of each variable on mode choice prediction, we analyze the global SHAP values from both the MNL and NN models. The corresponding SHAP values are visualized in Fig 3.

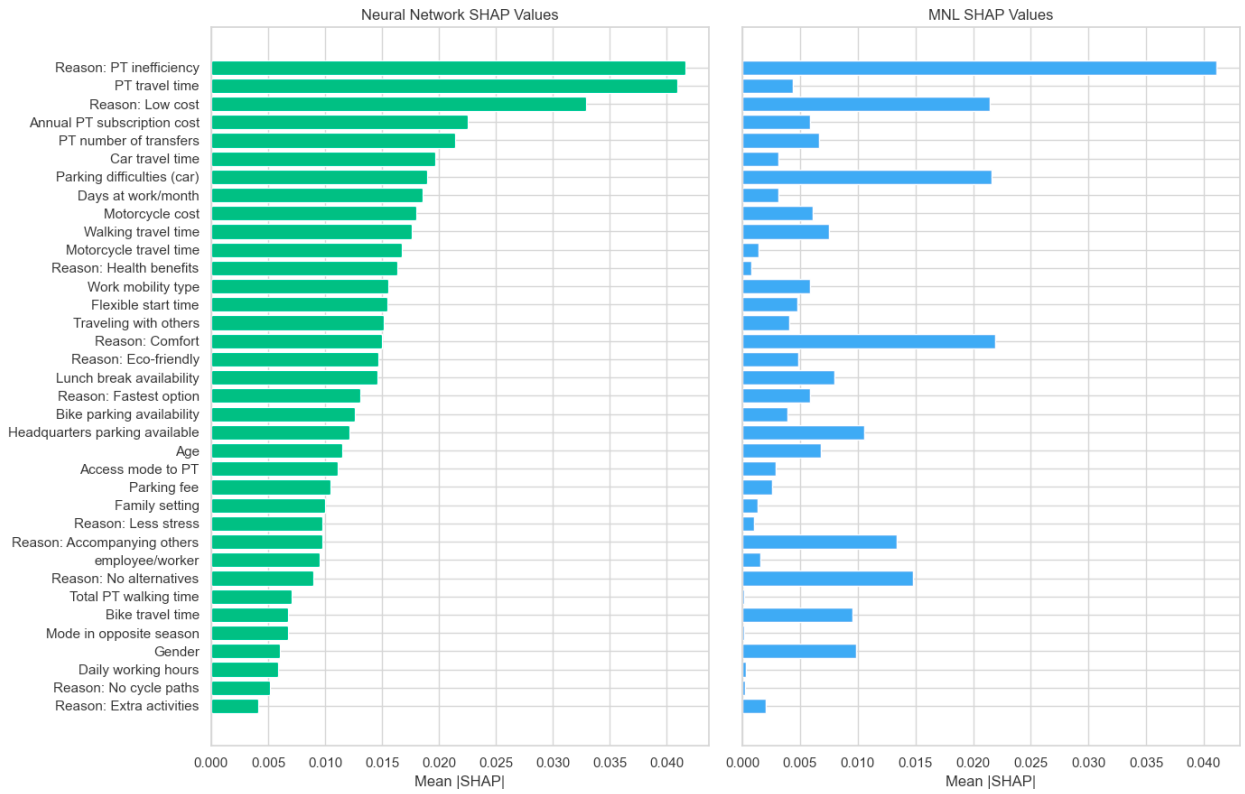


Fig 3: Global SHAP values for (a) NN and (b) MNL model

To facilitate a direct comparison between model structures, SHAP values from the NN and MNL models are plotted using a common feature ranking based on descending importance in the NN. This approach allows identification of both agreements and divergences in feature attributions. Several features exhibit consistently high importance across both models, notably “Reason: PT inefficiency,” “Reason: cost,” and “Parking difficulties (car),” indicating a shared understanding of the disutility associated with poor public transport quality, economic concerns, and car-related constraints. Likewise, both models assign low importance to features such as “Reason: No cycle paths,” “Reason: Extra activities,” and “Daily working hours,” confirming their limited contribution to explaining mode choice.

However, key differences emerge in how each model distributes importance among other features. The NN model assigns relatively high SHAP values to “PT travel time,” “Annual PT subscription cost,” “PT number of transfers,” “Car travel time,” and “Days at work/month,” suggesting that it captures both level-of-service and contextual workload factors more comprehensively. These variables are present in the MNL but with notably lower SHAP values, indicating that their marginal effects are either weaker or more diluted in the linear specification. In contrast, the MNL model places disproportionately high importance on variables like “Reason: Comfort,” “Reason: No alternatives,” and “Reason: Accompanying others,” which are ranked substantially lower in the NN model. This pattern likely reflects the MNL’s emphasis on a few dominant variables with strong isolated coefficients.

Another structural difference lies in the distribution of SHAP values: the MNL exhibits a sharp concentration of importance in a few features, while the NN spreads influence more evenly across a broader set. This reflects the NN’s capacity to account for non-linear interactions and shared variance, whereas the MNL focuses on additive, separable effects. Together, these findings demonstrate how SHAP can serve as a valuable diagnostic tool, revealing not only which features matter but also how model architecture shapes behavioral interpretation.

From a policy perspective, these differences in model behaviour are critical. The MNL model offers transparency and clear prioritization of a few dominant factors, which makes it highly suitable for interpreting user behavior and designing targeted interventions. In contrast, the neural network captures a broader range of interacting and attitudinal variables, revealing subtler influences that may be overlooked in linear models. This flexibility is valuable for identifying emerging or latent behavioural patterns.

3.3. Sustainable Mobility Insights

To encourage a shift from cars and motorcycles to more sustainable transport modes, it is essential to understand how commuters respond to changes in travel-related attributes. Elasticity values quantify how sensitive mode choice probabilities are to marginal changes in individual features (Ben-Akiva & Lerman, 1985), typically expressed as the percentage change in choice probability resulting from a 1% change in a variable. Table 2 presents the elasticity estimates for key variables from both the MNL and NN models, offering valuable insights for transportation policy. While both models identify car travel time as a negative factor, its elasticity in the MNL model (-0.025) is very small, indicating low behavioural sensitivity to travel time. In contrast, the NN model shows a stronger effect (-0.457), suggesting that the non-linear structure captures its importance better. The MNL model attributes high negative elasticity to car cost (-0.85) and perceived parking difficulty (-0.78), suggesting strong behavioral sensitivity to these monetary and spatial constraints. The NN model reflects similar directional influence but with reduced magnitudes (-0.51 for parking difficulty and -0.46 for car travel time), which is characteristic of its non-linear learning structure.

However, some discrepancies arise between the models. For example, “car cost” exhibits a negative elasticity in MNL (-0.85) but appears weak and even positive in the NN model ($+0.04$), indicating limitations of NN in capturing direct price sensitivity. Additionally, “parking fee” shows only a minor impact in both models (-0.03 and 0.05), despite being a frequently discussed policy lever. This suggests that commuters who regularly use cars may remain committed to this mode even when faced with rising parking costs, indicating low price sensitivity for this group.

For PT, both models identify the number of transfers and fare cost as primary deterrents. The MNL model shows stronger elasticities (-0.95 for transfers and -0.30 for fare), while the NN model supports these findings with more moderate values (-0.63 and -0.20 , respectively). These results validate known user preferences for direct, affordable public transport options. While environmental and health concerns do have negative elasticities in the MNL model (-0.11 and -0.01), they are either weak or reversed in the NN model, again illustrating how NN may suppress the effect of variables that are less frequent or interact weakly with others.

The identified sensitivity to parking-related costs and public transport attributes suggests that targeted policies, such as parking restrictions, fare subsidies, and improved transit connectivity, could effectively shift commuter behaviour toward more sustainable modes. These insights reinforce the value of data-driven tools for designing strategies that support long-term urban sustainability goals.

Overall, this analysis reveals that while both models agree on major behavioral drivers, the MNL offers clearer, more interpretable guidance for policy, especially when targeting pricing or structural interventions. The differences further justify the development of hybrid models, which integrate the interpretability of econometric frameworks with the flexibility and non-linear capabilities of neural networks, offering a more holistic understanding of commuter behavior for effective sustainable mobility planning.

Table 2: Comparison of average elasticity values for selected variables from MNL and NN models to evaluate their impact on mode choice

Variable	Elasticity (MNL)	Elasticity (NN)
Parking Fee on Car	-0.0334	0.0548
Car Cost on Car	-0.8516	0.0369
Time Driving on Car	-0.0257	-0.4571
Parking Difficulty Reason on Car	-0.7820	-0.5120
No. of Transfers on PT	-0.9460	-0.6347
Environmental Concerns on Car	-0.1132	-0.0384
Health Concerns on Car	-0.0128	0.0014
PT Fare on PT	-0.2981	-0.1986

4. Conclusion

This study examined the performance of MNL and NN models using commuter survey data from Rome to understand mode choice behavior and inform sustainable mobility strategies. A paired t-test showed that the differences between the results of the prediction of the two models are not statistically significant, indicating similar classification capabilities across both models. Despite comparable predictive performance, the MNL model offers clearer interpretability, enabling precise elasticity estimates and revealing strong behavioral sensitivities to travel cost, parking difficulty, and public transport transfers. These characteristics make MNL particularly suitable for policy analysis and the design of targeted interventions. The NN model complements this by uncovering complex, non-linear relationships that may be obscured in linear models, but lacks the same level of transparency for decision-makers.

The study identifies critical barriers to sustainable commuting in Rome, including high car dependency, insufficient public transport integration, and underdeveloped active transport infrastructure. Policy recommendations such as reducing transfer burdens, adjusting fare structures, and limiting parking accessibility are supported by both models. Importantly, the analysis highlights the value of combining econometric rigor with machine learning flexibility. Future work should explore hybrid modeling frameworks to leverage the strengths of both approaches, offering comprehensive insights for sustainable urban transport planning.

References

- Afsari, M., Varghese, K. K., Miristice, L. M. B., & Gentile, G. (2025). Comparing Choice Model and Machine Learning for Commuter Behavior Prediction: Evaluating Model Transferability Across Urban and Cultural Contexts. 2025 9th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), 1–6. <https://doi.org/10.1109/MT-ITS68460.2025.11223606>
- Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete Choice Analysis*. MIT Press.
- Bierlaire, M. (2025). Biogeme: A Free Software for Discrete Choice Modeling. <https://biogeme.epfl.ch/>
- Buchanan, J., & Walton, D. (2012). Motorcycle speeds at urban intersections. *Injury Prevention*, 18(Suppl 1), A191–A191. <https://doi.org/10.1136/injuryprev-2012-040590s.7>
- Eldafrawi, M., Varghese, K. K., Afsari, M., Babapour-dijojin, M., & Gentile, G. (2024). Machine Learning-Enhanced Conformal Prediction Approach for Road Traffic Accident Severity Assessment: A Case Study of Rome. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4679159>
- García-García, J. C., García-Ródenas, R., López-Gómez, J. A., & Martín-Baos, J. Á. (2022). A comparative study of machine learning, deep neural networks and random utility maximization models for travel mode choice modelling. 24th Euro Working Group on Transportation Meeting, 62, 374–382. <https://doi.org/10.1016/j.trpro.2022.02.047>
- 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>
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, 4765–4774.
- M. Afsari, M. Dastmard, L. M. Bresciani Miristice, & G. Gentile. (2024). Utilizing P-median and Machine Learning for Bike-Sharing Demand Prediction. 2024 IEEE International Conference on Environment and Electrical Engineering and 2024 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), 1–5. <https://doi.org/10.1109/EEEIC/ICPSEurope61470.2024.10751398>
- McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105–142). Academic Press.
- Shapley, L. S. (1953). A Value for n-Person Games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the Theory of Games (Vol. 2, pp. 307–317)*. Princeton University Press.
- Tabatabaei, M., Khavas, R. G., & Kashani, A. T. (2023). Feasibility of Estimating Macroscopic Fundamental Diagrams in Iran Traffic Network. *Proceedings of the 13th International Congress on Civil Engineering (13ICCE)*. <https://civilica.com/doc/1853127/>