International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(16s) 2026 pp. 974-987
https://doi.org/10.70917/ijcisim-2026-4625
Article

Analysing Commuters' Transit Mode Chosen Pattern in Urban Cities to Build Sustainable Intelligent Transit Mobility

J Sekaran [1], M Rameshkumar [2]

[1]PG & Research Dept. of Computer Science, Govt. Arts College (Autonomous), Nandanam, Chennai, India. Email: jogheesekaran@gmail.com
[2]PG & Research Dept. of Computer Science, Govt. Arts College (Autonomous), Nandanam, Chennai, India. Email: proframeshkumar@gmail.com

Abstract: Brisk urbanization and growing transportation demand have deepened the essential for sustainable and intelligent transit systems. Understanding commuter travel behaviour is important for building active transportation plans and mobility solutions. This research examines the commuter mode-choice behaviour using advanced approach based on the Random Forest algorithm. Here, travel dataset containing of 544 observations and 44 descriptive variables was used to analyse the aspects dominating the chosen of among private motorized transport, public motorized transport, and non-motorized transport modes. The Random Forest algorithm have attained an overall prediction accuracy of 89.02%, demonstrating superior capability in identifying complex nonlinear relationships among socio-economic, demographic, and travel-related variables. Feature importance analysis expressed that specified mode priority, automobile requirement, travel cost, driving license ownership, and travel distance meaningfully influence the commuter decisions. The findings give valuable insights for urban organisers and policymakers in scheming sustainable mobility plans and intelligent transportation systems possible of plummeting the congestion, emissions, and dependence on private vehicles.
Keywords: NA

1. INTRODUCTION: STUDY BACKGROUND

Urban transportation systems across the world are facing extraordinary pressure due to speedy population progress, growing vehicle ownership, and the nonstop development of metropolitan cities ([30], Sci. Rep., 2024). The increasing request for mobility has placed noteworthy strain on transportation infrastructure, resulting in heavy traffic congestion, lengthy travel times, enlarged delays, and reducing transportation performance. These factors are further strengthen by mounting greenhouse gas emissions, air pollution, noise pollution, and additional fuel consumption created by incompetent traffic conditions and extended vehicle wasting. Those problems not only surge transportation expenses, but also unfavorably disturb the economic output, environmental sustainability, and the complete excellence of urban life. Speaking those challenges demands the growth of sustainable and intelligent transportation systems reinforced by all-inclusive understanding of commuter travel behavior. Travel mode option is dominated by a mixture of socioeconomic, demographic, household, accessibility, and trip-related variables including income, vehicle ownership, travel time, travel cost, comfort, safety, and service reliability ([29], Sustainability, 2023). Understanding how those variables form the commuters' decisions allows transportation planners and policymakers to plan actual approaches which motivates sustainable travel modes, increase public transport services, reduce private vehicle dependency, and boost urban mobility ([38]).

Traditional transportation mode-choice research has fundamentally depended on discrete choice models such as the Multinomial Logit (MNL) and Nested Logit models, which are grounded on economic usefulness theory and offer interpretable assessments of the effects of descriptive variables on travel behaviour ([33], Sarkar Shraban, S., Basaruddin, K. S., & Islam, M. A. 2025). However, those approaches frequently depend on assumptions such as linear relationships and the liberation of unrelated replacements, restrictive their capacity to capture the complex, nonlinear communications present in practical world travel choices ([6], Bishop, C. M. (2006). Modern advances in machine learning and data mining have offered many comfortable and exact models for examining large-scale transportation datasets without imposing firm statistical assumptions. Amid those methods, the Random Forest algorithm has arisen as an actual collaborative learning technique ([1], Breiman, L. 2001). because of its high foretelling accuracy, strength against overfitting, capacity to execute high-dimensional data, and fitness to specify the qualified position of influential variables. Accordingly, this research pays a Random Forest classification model to evaluate commuters' behavioural patterns in urban transit mode choosing using travel survey data. The study focus to build an exact predictive structure for categorising travel modes while specifying the main factors dominating commuter behaviour. The results are anticipated to give appreciated decision-support data for policymakers, transportation officers, and urban developers in planning sustainable, capable, and intelligent transit systems that encourage environmental welcoming mobility and develop urban transportation amenities ([17], Kashifi, M. T., et al. 2022).

STUDY OBJECTIVES

1. To examine commuter mode-selection pattern using travel survey data. Research concentrated on investigating the attributes of commuters and specifying associations between socioeconomic, demographic, and trip-oriented parameters which dominates transportation mode priorities.
2. To develop a Random Forest model for predicting transportation mode selection. The study seeks to construct, train, and evaluate a Random Forest classification model capable of accurately predicting commuters' preferred transportation modes based on multiple explanatory variables.
3. To identify the most influential factors affecting commuter decisions. Using the variable importance measures generated by the Random Forest algorithm, the research aims to determine which factors contribute most significantly to travel mode selection and commuter behaviour.
4. To provide recommendations for sustainable and intelligent transit systems. Based on the analytical findings and predictive results, the study aims to propose practical recommendations that can support transportation policymakers and urban planners in developing efficient, environmentally sustainable, and data-driven transportation strategies that enhance urban mobility and improve the overall performance of public transit systems

STUDY CONTRIBUTIONS

This research contributes to the area of sustainable urban transit mobility by carefully examining commuters' transit mode selection patterns by applying the Random Forest algorithm and vast machine learning approaches which has capacity to capture the complex, non-linear associations among travel behaviour parameters. Unlike conventional statistical methods, currently, proposed approach specifies the relative reputation of socioeconomic, demographic, trip-oriented, and service excellence aspects which dominate commuters' mode selection decisions with greater projecting accuracy. This research delivers a data-driven outline for understanding transit mobility pattern, permitting transport officers and strategy deciders to build intelligent transit mode policies which promote the commuters to move from private transport to public transport and multimodal transport. Besides, this study proves that applicability of machine learning in travel required analysis and decision support, providing the appreciated insights for increasing the transit service preparation, reducing traffic bottleneck, greenhouse gas emissions, encouraging environmentally sustainable urban transportation.

2. LITERATURE REVIEW & RESEARCH GAP

Fastest growth in urban cities forced surely shifted travel requirements across urban cities worldwide ([30], Shafique, M. A., & Hato, E. 2021). Since, urban areas continue to enlarge connection with population, economic movements, and geographical limitations, So, demand of the competent transit systems has amplified significantly ([38], Diallo, et al. 2022). Urban population has caused in larger drive of commuters from home to office and vice versa, home to educational institutions and vice versa. Therefore, transit mode is demanded to adjust greater amount of daily commuters, ensuring availability, competence, and consistency. Even though, in several urban cities, improvement in transportation amenities not fit to current situation which generating considerable challenges for urban mobility ([5], James, G., Witten, D., Hastie, T., & Tibshirani, R. 2021).

Growth in connection with population, cumulative automobile ownership, and incessant urban development have deepened traffic bottleneck, environmental contamination, and fuel feasting. An increasing quantity of private automobiles has engaged significant burden on current road arrangements, causing in regular traffic waits, more travel times, and condensed transportation performance. Jammed traffic situation also contribute to greater fuel drinking due to lengthy automobile waiting and stop-and-go driving designs ([38], Diallo, et al. 2022). Furthermore, amplified vehicles usage creates notable emanations of greenhouse gases and extra air contaminates, leading to worsening air excellence, environmental dilapidation, and hostile public well-being collisions. Those factors have stimulated governments and transportation officers to pursue sustainable transportation answers which control reliance on private automobiles while encouraging greater resourceful and environmentally welcoming transit modes.

Understanding commuters' transit mode selection is basically depending on many attributes. Transit mode selection prefers decision-making practice through which every commutes choose a specific transit mode like private cars, public buses, metro trains, motorcycles, bicycles, or walking, for a specific trip. Those conclusion is dominated by various interconnected factors, including socioeconomic characteristics, demographic parameters, travel time, travel expenses, transit availability, service consistency, suitability, comfort, safety, and commuter’s priorities ([2], Breiman, L. 1996). Above said parameters interrelated in complex ways, examining commuter attributes is important for forecasting travel requirements, developing transportation facilities, and framing actual planning which promote the acceptance of sustainable transportation modes. A detailed understanding of travel mode selection ensures the officers to enhance public transportation systems, ease traffic bottleneck, increase drive performance, and help environmentally sustainable urban improvement ([39], Breyer, et al. 2021).

Conventional transit mode selection researches have normally depend on discrete choice models, specifically the Multinomial Logit (MNL) and Nested Logit models, to clarify commuter behaviour ([21], Molnar, C. 2022). Those approaches are grounded on utility maximization theory and estimate the chances of choosing a particular transportation mode according to the observed usefulness related with individual existing alternative ([31], Andersson, A., Kristoffersson, I., Rydergren, C., & Börjesson, M. 2025). Their theoretical basics and easiness of explanation have made them extensively valid in transportation investigation and preparation. However, those traditional methods naturally assume linear relationships among explanatory parameters and rely on statistical assumptions which may not exactly denote the difficulty of real-world travel behaviour ([24], Friedman, J. H. 2001). In habit, commuter conclusions are dominated by nonlinear connections, unseen dealings, and manifold inter-reliant parameters which are tough to capture using conventional statistical methods.

This research is based in Random Utility Theory (RUT), the Theory of Planned Behavior (TPB), and Activity-Based Travel Theory (ABTT), which collectively give detailed theoretical outline for understanding commuter travel mode choices ([31], Andersson, A., Kristoffersson, I., Rydergren, C., & Börjesson, M. 2025). Random Utility Theory disclosed that each commuter mode selection in transportation offers the highest observed effectiveness based on parameters like travel time, travel cost, convenience, comfort, and availability. Even though researchers unable to detect each parameters dominating every commuter's decision. This theory assumes that transit mode selection containing both observable and unobservable attributes. The Theory of Planned Behaviour quotes this perspective by insisting that travel decisions are not derived exclusively by economic reasons, but, also, are decided by behavioural aims, individual thoughts, particular standards, and observed behavioural control. This approach, enlighten the significance of psychological and social variables in designing commuters' transportation prioritises and decision-making manners ([23]), Chen, T., & Guestrin, C. 2016).

Activity-Based Travel Theory additionally spreads the reality of travel pattern by identifying the transit mode selection is a derived from the requirement of commuters to involve in regularly in their day to day works like job, education, shopping, hospital, and hobby. As per this approach, each person form their transit based on the travel timing, destination, and series of activities rather than deciding the remote travel choices. Subsequently, travel mode choice is dominated by the communication amid activity timetables, convenience, family personalities, and transportation facilities. Jointly, those three theoretical viewpoints elucidate that travel mode priorities rely on a grouping of economic, behavioural, psychological, and availability features ([33], Sarkar Shraban, S., Basaruddin, K. S., & Islam, M. A. 2025). Through inter-connecting those additional methods, this research founds a solid theoretical basis for examining commuter's behaviour and supports the use of machine learning methods, especially, the Random Forest classification model, to specify the primary factors dominating the urban transportation mode selection.

Traditional mode choice approaches have practiced in transportation examination to examine and forecast commuters' travel pattern ([12], Ben-Akiva, M., & Lerman, S. R. 1985). Among the frequently applied techniques are the Multinomial Logit (MNL)**,** Nested Logit (NL), and Mixed Logit (ML) techniques. Those techniques are built on the values of random advantage expansion, where commuters are expected to choose the transit mode which give the maximum observed usefulness amid existing possibilities. Their solid theoretical basis ensures scholars to estimate the dominance of descriptive parameters like traveling time, travel expenses, monthly income, vehicle ownership, availability, and service excellence on transportation mode selection. Due to their statistical interpretability, those approaches have been widely used to assist transportation preparations, policy assessments, requirement forecasting, and infrastructure improvement ([5], James, G., Witten, D., Hastie, T., & Tibshirani, R. 2021).

Despite their extensive implementation and appreciated pattern understandings, conventional transit mode selection approaches have numerous constraints when implemented to multifaceted and high dimensional transportation datasets. Those techniques normally depend on assumptions concerning linear associations, predefined functional procedures and particular statistical attributes which may not sufficiently signify real-world travel pattern. When number of explanatory parameters rises and connections among factors reduce the forecasting efficiency of those approaches. In specific, they frequently difficult to capture nonlinear associations, unknown designs, and complicated communication among socioeconomic, demographic, and trip-related attributes ([25], Kuhn, M., & Johnson, K. 2019). Those constraints normally promoted the researchers to discover innovative machine learning methods which can handle the complex data structures more successfully while increasing forecasting accuracy and assisting told transportation strategies and decision-making.

Machine learning techniques such as Random Forest, Gradient Boosting, Support Vector Machines (SVM), and Neural Networks have increasingly been applied to transportation research due to their ability to model complex and nonlinear relationships within large datasets ([4], Hastie, T., Tibshirani, R., & Friedman, J. 2009). These algorithms are widely used for travel demand prediction, traffic flow forecasting, congestion analysis, route optimization, and intelligent transportation systems (ITS) ([38], Diallo, et al. 2022). Unlike traditional statistical approaches, machine learning models can automatically learn patterns from historical and real-time transportation data, enabling more accurate predictions and adaptive decision-making. Their capability to process diverse data sources, including traffic sensors, GPS traces, mobile applications, and smart infrastructure, has significantly enhanced transportation planning and management. Machine learning techniques plays an important part in increasing the competence and dependability of current transportation structure ([24], Friedman, J. H. 2001). Approaches such as Random Forest and Gradient Boosting has vast forecasting ability by managing high-dimensional datasets and capturing involved connections among parameters, while promoting Vector Machines are active for classification and regression duties linking complex traffic designs. Neural Networks, especially deep learning approaches, have established significant performance in investigating large-scale transportation data, containing traffic bottleneck detection, travel time assessment, and event forecast. Since, transportation in urban cities continue to produce more amount of data, machine learning has become an important model for building intelligent, data-driven resolutions which support sustainable mobility, decrease traffic bottleneck, and increase overall transportation network performance ([26], Kuhn, M., & Johnson, K. 2013).

Available researches in transit mode selection need estimate which mainly concentrated on building the predictive accuracy of machine learning approaches, frequently insisting performance metrics like accuracy, precision, recall, and error decline ([25], Kuhn, M., & Johnson, K. 2019). Apart from that, those researches prove the competence of latest algorithms in predicting travel pattern, they often give limited vision into the causing parameters which dominates distinct travel mode selections. As a result, the explanation of machine learning approaches remains vital challenge, creating more barriers for scholars and strategy officers to explain how socioeconomic variables, travel variables, and behavioral variables contribute in taking transit mode selection decisions. Since, having absence of interpretability lessen the real-world worth of predictive models for assisting evidence-based plans and policy formulation.

Building an applicable forecast method, this study specifies the primary pattern determinants which dominates commuter's mode selection through feature significance inquiry, thereby improving the interpretability of the model. The outputs contribute to build sustainable transportation policies by giving valuable insights into the aspects disturbing travel pattern which surly assist the plan of principles that promote competent, environmentally welcoming, and available transportation systems ([19], Hagenauer, J., & Helbich, M. 2017).

3. METHODOLOGY: PREVIOUS STUDY

The TRIP (2025) research which taken as base paper suggested that combined structures of Discrete Choice Models (DCM) with Machine Learning (ML) techniques assisted to obtain the transit mode behaviour estimation. The research attained the 60% overall accuracy, 58% precision and 59% recall which representing that combining behavioural models with data-driven models can increase the forecast performance compared with conventional approaches alone. Even though, the forecasting aptitude remains reasonable, specifying chances for future improvements. This research pinpoints the possible of hybrid DCM+ML models in taking both traveller decision-making behaviour and complex nonlinear relationships in transport data. For forthcoming research, advanced algorithms such as LightGBM, XGBoost, and Explainable AI (XAI) techniques can be incorporated to increase the forecast accuracy, interpretability, and generalizability for clever transportation plans.

Study	Overall Accuracy (%)	Precision (%)	Recall (%)
TRIP (2025)	60	58	59

Table 1 - Classification Report of TRIP research

According to explanation provided here, this study used the hybrid Discrete Choice Model (DCM) along with Machine Learning (ML) algorithms for predicting the travel behaviour of commuters. Even though, the perfect machine learning algorithm (e.g., Random Forest, SVM, Neural Network, XGBoost) is not used in this research. The explanations only telling that the research combines DCM with ML methods and proposes forthcoming research using LightGBM, XGBoost, and Explainable AI (XAI). Those specifies that the combined DCM and ML methods attained restrained forecasting capacity. Suppose building the demonstration of both commuters decision-making attitudes and nonlinear associations. This researcher suggested incorporating many advanced algorithms such as LightGBM, XGBoost, and Explainable AI (XAI) in later researches increase the estimated accuracy, interpretability, and generalizability.

3.1 PROPOSED STUDY: RESEARCH STRUCTURE

The methodological structure implemented in this research containing the following steps.

1. Data gatherings and Pre-processing.

2. Feature engineering and Data cleaning.
3. Random Forest approach implementation.
4. Performance Assessments.
5. Forthcoming Vital Examinations.
6. Strategy explanations and suggestions.

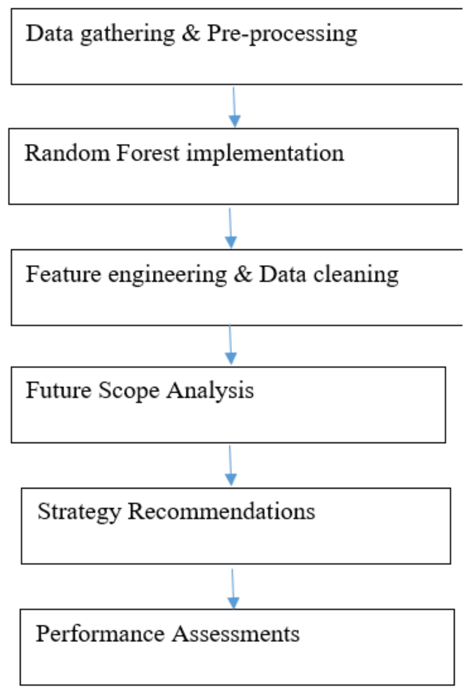

Fig 1- Architecture of work flow

3.2 Dataset Description

Dataset used in this research contains 544 observations and 44 variables, generally applied in transportation study. Those 44 attributes are belonged to different category like Socio-demographic factors, Household characteristics, Travel characteristics, Accessibility characteristics, Built environment / Urban form characteristics, Traffic and Transport network characteristics, Survey or Identification variables. Here, ModeRP variable is fixed as target variable.

In this dataset, Commuter observations have gathered by conducting the ample travel behavior survey. Each observation denotes an individual commuter and captures data connected to their travel features and mode-choice behavior. This survey was planned to collect data to illustrate mode selection patterns, allowing the analysis of issues which dominates the transit decisions in an urban background. The gathered dataset helps as the main source of information for planning and assessing the machine learning model used in this study. Attributes which involved in the dataset belongs different categories like the demographic, socio-economic, transportation, and accessibility parameters of the commuters living in urban cities. Those attributes such as age, gender, income, occupation, education, household variables, travel time, travel cost, trip aim, vehicle ownership, and availability of the transit options together offer the detailed picture of the reality which may dominates the commuters' transit choose option. The variety of parameters permits the Random Forest model to specify the complex associations among different elements, assisting precise forecast and meaningful understanding of travel behavior.

3.3 Target Variables

The dependent variable is revealed preference mode choice (ModeRP), categorized as:

Class 1 – Private Motorized Transport

Class 2 – Public Transport

Class 3 – Non-Motorized Transport

3.4 Random Forest Algorithm

Random Forest is the collective machine learning algorithm containing many or multiple decision trees which works jointly to improve the forecasting correctness and model steadiness. Individual decision tree is skilled autonomously using a random subset of observations produced through bootstrap sampling, whereas any random subset of predictor variables is chosen at each split during the tree-building progression. Both mixture of data sampling and random feature selection encourages variety among the distinct trees, reducing the risk of overfitting and increasing the model's capacity to generalize to hidden data. The final forecast is accessed by combining the result of all decision trees through mainstream voting for grouping the aims or averaging for regression tasks. Due to its heftiness, capacity to handle high-dimensional datasets, and competence to model complex nonlinear associations, Random Forest is being one of the maximum extensively used machine learning algorithms for predictive study across different application fields, including transportation research.

4. RESULTS AND DISCUSIONS

The final forecast is attained through majority voting:

RF(x) = Majority Vote {T1(x), T2(x), …, Tn(x)}

where: RF(x) = Random Forest prediction, Ti(x) = Prediction from individual decision tree, n = Number of trees

Evaluation Metrics

The model performance is evaluated using the following formulas.

$$\boldsymbol{Accuracy} = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})} * 100$$

Where, TP (True Positive): Exactly forecasted positive cases. TN (True Negative): Properly forecasted negative cases. FP (False Positive): Wrongly projected positive cases. FN (False Negative): Incorrectly forecasted negative cases.

$$\mathbf{Precision} = \frac{TP}{(\mathrm{TP} + \mathrm{FP})} * 100$$

Here, TP (True Positives): Exactly forecasted positive cases. FP (False Positives): Wrongly projected positive cases.

$$\mathbf{Recall} = \frac{TP}{(\mathrm{TP} + \mathrm{FN})} * 100$$

Here, TP (True Positives): Exactly forecasted positive cases. FN (False Negatives): Positive cases which this model erroneously projected as negative.

F1-Score: F1 = 2 × (Precision × Recall)/ (Precision + Recall), here, TP, TN, FP, and FN denote usual classification results.

Dataset size: 544 observations, 44 Variables

Training samples (70% split): Approximately 380 observations

Testing samples (30% split): Approximately 164

observations

Number of trees created: 200

Classification Report: **Accuracy = 89.02%**

Class	Precision	Recall	f1-score	support
1 (Private)	0.953846	0.925373	0.939394	67.000000
2 (Public)	0.837500	0.943662	0.887417	71.000000
3 (Non- Motorized)	0.894737	0.653846	0.755556	26.000000
accuracy	0.890244	0.890244	0.890244	0.890244
macro avg	0.895361	0.840960	0.860789	164.000000
weighted avg	0.894106	0.890244	0.887747	164.000000

Table 2 - Classification Report

Using the Random Forest algorithm, study attained an overall classification accuracy of 89.02%, signifying that it exactly classified the majority of transportation mode cases. This \method confirmed solid predictive performance across the three classes, with a weighted precision of 89.41%**,** weighted recall of 89.02%**,** and a weighted F1-score of 88.77%, echoing balanced overall performance despite variances in class allocation. The macro-average precision (89.54%)**,** recall (84.10%)**,** and F1-score (86.08%) imply that the classifier achieved steadily across three classes, even though, few variant survives in its capability to specify every transit mode. Those outcomes specify that the Random Forest algorithm is applicable for multiclass transportation mode classification.

A nearer analysis of the class wise performance displays that the method attained finest for the Private transportation class, attaining a precision of 95.38%, recall of 92.54%, and F1-score of 93.94%, signifying outstanding competence in precisely specifying private transport instances while reducing false positives. The Public transportation class also showed resilient performance, with a precision of 83.75%, recall of 94.37%, and F1-score of 88.74%, demonstrating that most public transport samples are precisely acknowledged, although, having more private or non-motorized samples may assist to get more accuracy. In contrast, the Non-Motorized class attained a precision of 89.47%, but, this reasonably lower recall of 65.38%**,** resulting in an F1-score of 75.56%. This proposes that method was highly accurate when forecasting non-motorized transit, it unsuccessful to specify a sizeable proportion of actual non-motorized instances, likely due to the smaller number of training samples for this class (support = 26) or coinciding attitudes with other transit nobilities. Overall, the conclusions proved that the Random Forest approach delivers unfailing classification performance with further betterment wanted to improve the discovery of the minority non-motorized class.

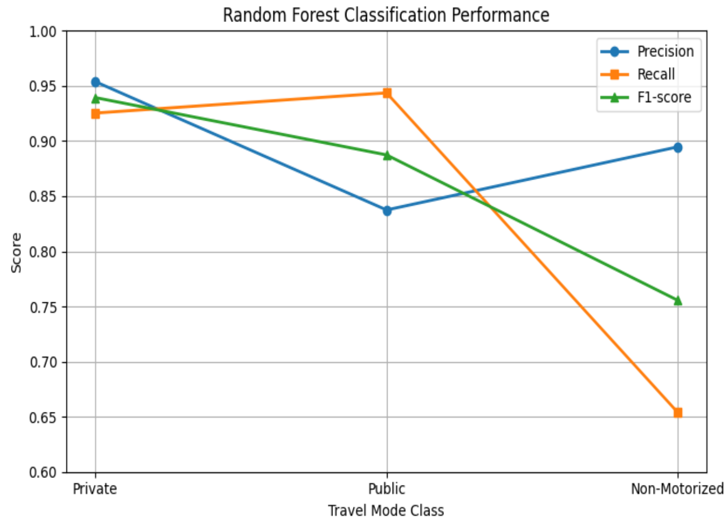

Fig 2 – Random Forest Classification Performance

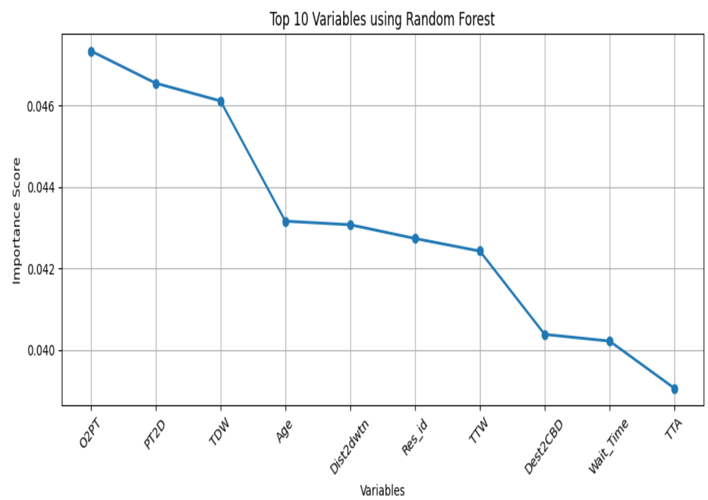

Fig 3 – Top Ten Variables Using Random Forest

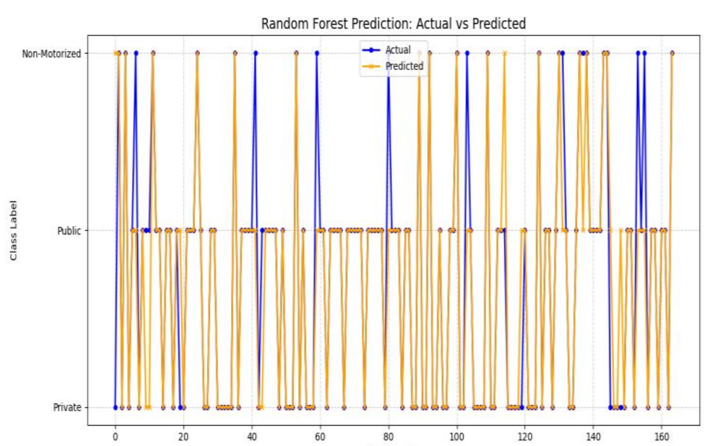

Fig 4 – Random Forest Prediction:Actual vs Predicted

Dataset Shape: (544, 44)

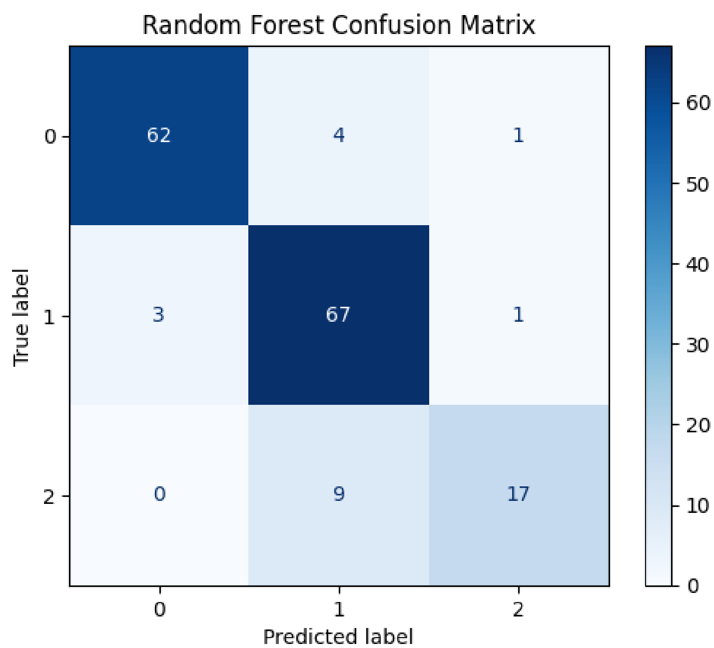

Fig 5 – Random Forest Confusion Matrix

Rank	**Variable**	**Interpretation**
1	ModeSP	Stated Preference Mode Choice
2	Auto_nec	Automobile Necessity
3	Exp_val	Expenditure Value
4	DLO	Distance to Local Opportunity (or Destination Location)
5	ODDEucl	Origin–Destination Euclidean Distance
6	TDA	Travel Distance by Automobile
7	TDPT	Travel Distance by Public Transport
8	Inc_val	Income Value
9	Cost_car	Travel Cost by Car
10	TTA	Travel Time by Automobile

Table 3 - Top Ten Influential Variables & Interpretation

Variable Position	**Variable**	**Importance**

1	ModeSP	0.267674
36	Auto_nec	0.084084
10	Exp_val	0.051914
12	DLO	0.041820
30	ODDEucl	0.029044
23	TDA	0.028859
25	TDPT	0.028017
9	Inc_val	0.026425
33	Cost_car	0.025991
22	TTA	0.025690

Table 4 - Top 10 Feature Importance Variables

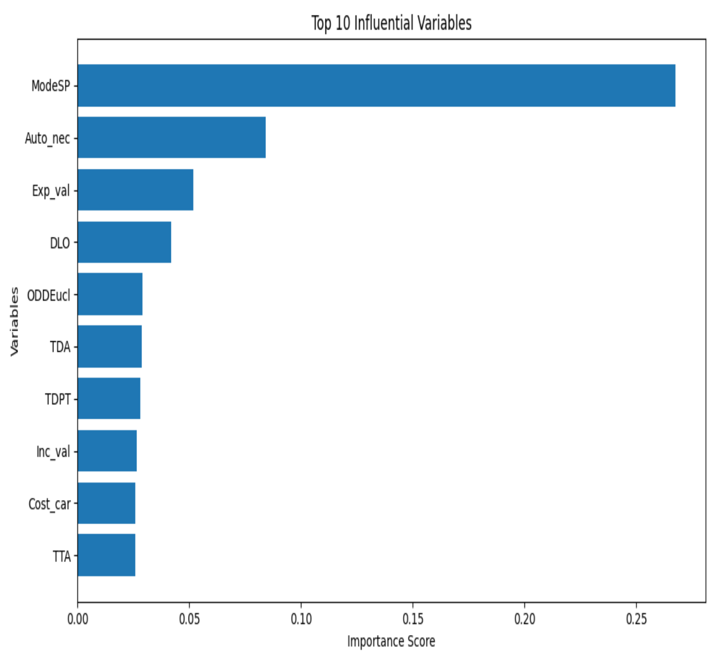

Fig 6 - Top Ten Influential Variables

The results proved that commuter transit mode choice is dominated by a blend of individual characteristics and transit mode parameters, indicating the multifaceted nature of travel mode choice in urban cities. Variables like age, gender, income, education, occupation, household alignment, vehicle ownership, and driving licence ownership outline commuters travel priorities, while transportation-linked variables including travel time, travel cost, service reliability, accessibility, comfort, safety, connectivity, and travel distance meaningfully disturb the attractiveness of various transit modes. The solid dominance of specified mode priorities additionally specifies that behavioural attitudes, individual insights, and reputable travel habits play the main role in commuter decision-making, whereas automobile reliance imitates the significance of private transit availability in deciding travel options. Furthermore, travel cost and travel distance effect commuters' assessment of existing transit choices, along with affordable, accessible, and useful services being more likely to inspire public transport acceptance. Those outcomes insist the requirement for transportation planning and strategy interventions that concurrently talk the behavioural parameters, increase public transport excellence, improve multimodal availability and last-mile connectivity, lessen the travel expenses and travel times, and encourage justifiable travel behaviour to help competent user-centred, and environmentally sustainable urban mobility systems.

5. CONCLUSION, LIMITATION, FUTURE WORK

The above research examined the travel behavior of commuters in choosing the transit mode using the Random Forest algorithm. Key parameters are identified which dominating the travel choices. This study reveals the how commuter's decision connected with collection of demographic, socioeconomic, transportation, and accessibility parameters in selecting the transit mode. Here, Random Forest algorithm established solid projecting capacity in dealing nonlinear connections amid parameters while preserving strength against overfitting, attaining an **overall accuracy of 89.02%**. The outcome ensures that improvement attained in terms of overall accuracy compared to TRIP research which shown in the table 1. Result of the model offers an unfailing and effective structure for forecasting commuter travel pattern and is exacted matched for transportation demand analysis, helping data-driven transportation preparation, strategy construction, and the building the intelligent and sustainable urban mobility arrangements. Feature importance analysis further specified mode priorities, automobile reliance, travel expenditure, driving license ownership, and convenience parameters become the greatest domination variables affecting commuters' transportation mode priorities, signifying the important part of behavioral, economic, and accessibility-related variables in deciding travel decisions. Result of this study further prove that machine learning techniques give appreciated decision-support abilities for sustainable transportation policies by exactly specifying the factors of commuter travel pattern and enlightening the travel demand prediction. The solid projecting performance of the Random Forest model highlights the possible of data-driven methods for transportation mode-choice analysis, allowing the transportation agencies and strategy makers to frame the evidence based policies which promotes the public transport practice, increase service quality, and encourage multimodal transportation choices. Furthermore, the identification of primary dominating parameters gives real-world direction for planning targeted interventions meant at sinking traffic bottleneck, dropping carbon releases, boosting transportation availability, and uplifting naturally sustainable travel behaviour. Subsequently, the suggested method contributes to the improvement of competent, strong, and smart urban transportation systems adept of supporting long-term sustainable mobility goals.

Despite achieving high overall classification performance, several limitations should be acknowledged. The model exhibited reduced effectiveness in identifying the Non-Motorized class, as evidenced by its relatively low recall, indicating that many instances of this class were misclassified. This limitation may be attributed to the smaller number of non-motorized samples compared with the private and public classes, resulting in class imbalance and reduced learning capability for minority classes. In addition, the Random Forest algorithm depends highly on the excellence and representativeness of the input features. Therefore, overlapping variables among transit modes may have contributed to classification faults. The assessment was also conducted on a limited dataset, which may limit the generalizability of the results to other geographic areas, travel patterns or data gathered situations.

Future study should concentrate on refining the classification efficiency of the marginal Non-Motorized class through methods like class balancing, oversampling approaches (e.g., SMOTE), or cost sensitive learning. Enlarging the dataset with additional varied and representative transportation observations from diverse areas and time periods would expand the robustness and generalizability of the methods. Furthermore, examining superior machine learning and deep learning algorithms, like Gradient Boosting, XGBoost, LightGBM, or neural network based methods may further improve classification precision. Future researches could also integrate more contextual features like weather conditions, traffic report, temporal patterns or GPS trajectory patterns to increase the model's capacity to differentiate between transportation methods with similar travel patterns.

ACKNOWLEDGEMENT

I would like to express my sincere thanks to Dr. M. Rameshkumar for his invaluable guidance, professional advice, and continuous support throughout this work.

REFERENCES

1. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.
2. Breiman, L. (1996). Bagging Predictors. Machine Learning, 24(2), 123–140.
3. Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
4. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
5. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning (2nd ed.). Springer.
6. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
7. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
8. Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press.
9. Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann.
10. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. (2016). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.
11. Ortúzar, J. de D., & Willumsen, L. G. (2011). Modelling Transport (4th ed.). Wiley.
12. Ben-Akiva, M., & Lerman, S. R. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press.
13. McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics.
14. Cheng, L., Chen, X., De Vos, J., Lai, X., & Witlox, F. (2019). Applying a random forest method approach to model travel mode choice behavior. Travel Behaviour and Society, 14, 1–10.
15. 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.
16. 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, 188, 116253.
17. Kashifi, M. T., et al. (2022). Predicting the travel mode choice with interpretable machine learning techniques: A comparative study. Travel Behaviour and Society.
18. Ali, N. F. M., et al. (2022). The identification of significant features towards travel mode choice and its prediction via optimised random forest classifier. Journal of Transport & Health, 25, 101362.
19. Hagenauer, J., & Helbich, M. (2017). A comparative study of machine learning classifiers for modeling travel mode choice. Expert Systems with Applications.
20. Ermagun, A., & Levinson, D. (2018). Public transit, active travel, and machine learning applications in travel behavior. (Relevant transportation research).
21. Molnar, C. (2022). Interpretable Machine Learning (2nd ed.).
22. Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems.
23. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of KDD.
24. Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics.
25. Kuhn, M., & Johnson, K. (2019). Feature Engineering and Selection. CRC Press.
26. Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer.
27. Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. Transportation Research Part D.
28. Ding, C., Wang, D., Liu, C., Zhang, Y., & Yang, J. (2017). Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance.
29. Rasouli, S., & Timmermans, H. (2014). Applications of machine learning in travel behaviour research.
30. Shafique, M. A., & Hato, E. (2021). Incorporating MNL Model into Random Forest for Travel Mode Detection. Mehran University Research Journal of Engineering and Technology.
31. Andersson, A., Kristoffersson, I., Rydergren, C., & Börjesson, M. (2025). Modelling travel mode choice on combined data sources using machine learning methods. Transportation Research Procedia.
32. Chauhan, R. S., et al. (2025). Causation versus prediction in travel mode choice modeling. npj Sustainable Mobility and Transport.

33. Sarkar Shraban, S., Basaruddin, K. S., & Islam, M. A. (2025). Machine learning in travel mode choice studies: A systematic literature review of applications, methods, and challenges. Results in Engineering.
34. Rajaure, T. (2025). Mode Choice Prediction: Comparing Econometric Models with Combination of Machine Learning Models. International Conference on Transportation and Development.
35. Konstantinou, C., Mantouka, E., Giouroukelis, M., & Vlahogianni, E. I. (2025). Comparing Random Forests and Multinomial Logit Models for Urban Travel Mode Choice Under Innovative Traffic Management Strategies. Lecture Notes in Mobility.
36. Martín-Baos, J. Á., López-Gómez, J. A., Rodríguez-Benítez, L., Hillel, T., & García-Ródenas, R. (2023). A prediction and behavioural analysis of machine learning methods for modelling travel mode choice. (Preprint).
37. Ghosh, T., & Nagaraj, N. (2024). Evaluating the Determinants of Mode Choice Using Statistical and Machine Learning Techniques in the Indian Megacity of Bengaluru. (Preprint).
38. Diallo, et al. (2022). Estimation of minority modes of transportation based on machine learning approach. Procedia Computer Science.
39. Breyer, et al. (2021). Travel mode classification of intercity trips using cellular network data. Transportation Research Procedia.
40. Ali, N. F. M., et al. (2022). Travel mode choice modeling using machine learning and feature importance analysis for active commuting. Journal of Transport & Health.