Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Predicting the Intention of Public and Private Vehicle Users to Shift to Public Transport Using Data Mining Techniques

Sk. Wasim Anwar [1], Ranja Bandyopadhyaya [2]
[1] *PhD Scholar, Department of Civil Engineering, National Institute of Technology Patna, Bihar, India. Email: sk.a.phd19.ce@nitp.ac.in*
[2] *Associate Professor, Department of Civil Engineering, National Institute of Technology Patna, India. Email: ranja@nitp.ac.in*

ARTICLE INFO

Received: 20 Dec 2024
Revised: 18 Feb 2025
Accepted: 26 Feb 2025

ABSTRACT

With increasing urbanization and the rising burden of private vehicle usage, encouraging a modal shift toward public transportation is essential for sustainable urban development. This study investigated the factors influencing the intention of commuters to shift to public transport by applying data mining techniques to behavioral and demographic data. A structured questionnaire collected data on perceptions of safety, comfort, time efficiency, convenience, and socioeconomic variables in the City of Kolkata, India. 345 questionnaires were considered, which were collected from public and private vehicle users. Using the WEKA tool, the PART decision tree classifier was employed to extract interpretable rules that explain commuter intentions. The study found that positive perceptions, travel convenience, and safety are important factors of future public transport usage, while youths having low income and adults show unwillingness unless quality conditions are satisfied. The study offers practical insights for transport policymakers to design targeted strategies that enhance service quality and influence commuter attitudes. The findings demonstrate the effectiveness of rule-based machine learning in predicting travel behavior and support the formulation of data-driven mobility policies.

Keywords: Public Transport, Private Vehicles, Mode Choice, Data Mining, WEKA, PART Classifier, Commuter Behavior, Decision Trees, Travel Intention.

INTRODUCTION

The transportation systems are essential to the sustainable growth of modern cities. For increasing urbanization, it is a emergent need to understand and influence the mobility choices of citizens, particularly their willingness to adopt public transport over private vehicles. Numerous studies have explored the behavioral, socioeconomic, and psychological factors that drive such choices. This section reviews relevant literature on transport mode choice behavior, data mining applications in transportation, and the use of decision tree models for analysis.

Travel behavior has been a central theme in transportation research, focusing on how individuals make decisions. Ben-Akiva and Lerman (1985) laid the foundation with their work on discrete choice models, highlighting that transport choices are influenced by socio-demographic factors, cost, travel time, and comfort. More recent studies, such as those by Schwanen et al. (2004) and Shiftan et al. (2008), have incorporated psychological constructs such as attitudes, perceptions, and travel satisfaction. Research introduced integrated choice and latent variable (ICLV) models, which combine choice modeling with latent psychological constructs such as attitude, perception, and satisfaction (Walker & Ben-Akiva, 2002; Vij & Walker, 2016). These models improved behavioral realism and enabled deeper insights into travel preferences. Yet, they still required strong assumptions about variable distribution and model specification

Factors like travel comfort, convenience, perceived safety, and trip purpose have been consistently identified as key determinants in transport mode choice. Jou et al. (2010) emphasized that income level and vehicle ownership significantly affects the likelihood of switching to public transport, while Susilo and Axhausen (2014) focused on habitual behavior and its resistance to change despite improvements in public transport systems.

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Redman et al. (2013) emphasized that service quality—specifically comfort, frequency, and travel time—is crucial for attracting car users to public transport. Similarly, Eboli and Mazzulla (2011) found that user satisfaction with comfort and safety significantly predicted continued use of public transport. Cao and Cao (2017) reported that individuals' perception of public transit's reliability and safety could outweigh objective attributes like cost or distance in influencing mode choice. Socio-demographic variables also play a significant role. Kamargianni and Polydoropoulou (2014) demonstrated that younger, lower-income individuals and students are more inclined to shift to public transport, provided that affordability and accessibility are ensured. Purpose of travel (e.g., commuting vs. leisure) has been shown to affect travel flexibility and openness to public modes (Gärling & Axhausen, 2003). Anwar & Bandyopadhyaya (2024) focused individual's belief about benefits of public transport like reduction in congestion and pollution and believe that their choice of use of public transport helps in reducing congestion and pollution.

With the advancement of intelligent transportation systems, techniques such as classification, clustering, and association rule mining have been widely used to uncover hidden patterns in transport datasets. Witten and Frank (2005) were among the first to apply machine learning tools like Weka to transportation data, demonstrating the feasibility of such approaches in predicting travel modes and preferences.

Zhong et al. (2016) used clustering and classification techniques to analyze smart card data and predict commuting patterns, while Jain et al. (2013) applied decision tree algorithms to study passenger behavior in Indian metropolitan cities. These methods offer the advantage of handling noisy, non-linear, and heterogeneous data effectively, thereby improving the accuracy of transport modeling and planning.

Decision trees are popular for their interpretability and ability to generate human-understandable rules. They are particularly useful in policy-making contexts where transparency in prediction logic is crucial. Studies such as Karlaftis and Vlahogianni (2011) and Yap et al. (2018) have employed decision tree classifiers (e.g., C4.5, J48, CART) to identify key variables affecting public transport use.

These models divide the data into subgroups based on attribute values and produce rules that explain decision outcomes. For instance, Ahmed et al. (2020) utilized decision trees to assess factors influencing bus ridership in Southeast Asia and found that convenience and travel time were primary determinants. Similarly, Kumar and Vanajakshi (2015) demonstrated how decision trees could be used to forecast mode shift behavior in mixed traffic conditions.

The Weka tool, an open-source platform for machine learning, has made the application of such models easily available to researchers and is widely used in transportation research for its open-source accessibility and comprehensive suite of machine learning algorithms (Hall et al., 2009). The PART classifier, in particular, combines the strengths of decision trees and rule-based classifiers to produce compact and interpretable models. Hall et al. (2009) noted that PART efficiently deals with categorical data and is robust against overfitting when dealing with medium-sized datasets such as those typical in transport surveys. Mohammad et al. (2017) employed PART to classify user satisfaction levels among public transport users. Zhou et al. (2019) applied it to develop travel mode classification models based on urban travel survey data, demonstrating the method's ability to generate actionable policies.

METHODOLOGY

The data mining method predicts the intention of public transport and private vehicle users to use public transport in the future. The process has been divided into three stages:

i. Stage I: Collection of data: The data was collected from different transport users, i.e., public transport and personal vehicle users. Twelve attributes were used from the collected data. The dataset was prepared and used for data mining, as mentioned in Table 1.

ii. Stage II: Processing of dataset: The dataset was classified using part rules using Weka tools. All the attributes were used in nominal types. A total of 345 instances were used for classification rules. The process starts with

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

matching the origin and ending points to build suitable rules to find the intentions of public transport and private vehicle users to use public transport in the coming days.

iii. Stage III: Develop a decision tree model and rules: A decision tree model is developed to predict the rules, which start from five nodes: travel comfort, convenience, perception, annual income, and purpose of travel. The twenty rules are built from the tree, which clearly shows the intentions of public and private vehicle users to use public transport in the future. In the result section top ten important rules were shown to avoid making the paper too long. Some nodes and leaves influence the users, and some do not.

DATA

The data was collected in Kolkata, with regular commuters of public transport and personal vehicles from public transport and private vehicle users. The data was collected by preparing a questionnaire, which includes demographic factors like age, gender, education level, annual family income, the purpose of travel and mode of travel, attitude towards perception, convenience, comfort, duration of travel, safety, and intention towards public transport. A total of 550 data were collected between November 2021 and October 2022 in Kolkata; 345 responses were considered, which included 233 public transport users and 112 personal vehicle users. Details of data are shown in Table 1.

Table1: Details of data

Attributes	Details of attributes	Instances	Frequency	Data in %
Age	Age	Young (<25 years)	43	*12%*
		Adult (25 to 45 years)	192	*56%*
		Middleage (45 to 60 years)	105	*30%*
		Old (> 60 years)	5	*1%*
Gender	Gender	Male	212	*61%*
		Female	133	*39%*
Educationlevel	Education level	Tenth	116	*34%*
		Twelth	92	*27%*
		Graduate	137	*40%*
Annualincome	Annual income of family	lowIncome (below 5 lakh)	340	*99%*
		highIncome (above 5 lakh)	5	*1%*
Purpose	Purpose of travel	NonWork	88	*26%*
		Work	257	*74%*
Mode	Mode of travel	PersonalVeh	112	*32%*
		SharedPT	233	*68%*
TravelPerception	My travelling to work on public transport is/ will be	Good	251	*73%*
		Neutral	87	*25%*
		Bad	7	*2%*
TravelConvenience	My travelling to work on public transport is/ will be	Convenient	252	*73%*
		Neutral	77	*22%*
		Inconvenient	16	*5%*

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Attributes	Details of attributes	Instances	Frequency	Data in %
TravelComfort	My travelling to work on public transport is/ will be	Comfortable	237	*69%*
		Neutral	90	*26%*
		Uncomfortable	18	*5%*
TravelTiming	My travelling to work on public transport is/ will be	TimeSaving	191	*55%*
		Neutral	118	*34%*
		TimeTaking	36	*10%*
TravelSafety	Use of public transport in the city is	Safe	226	*66%*
		Neutral	117	*34%*
		Unsafe	2	*1%*
UsePTinFuture	I intend to travel by public transport in the coming months	Yes	257	*74%*
		NotSure	88	*26%*

RESULTS

The data set was processed with 12 attributes and 345 instances using a data mining tool, i.e., Weka 3.8.6, to find the suitable rules to find the intention of using public transport in the future for public transport commuters and private vehicle commuters. The scheme was Weka.classifiers.rules.PART -C 0.25 -M 2 and test mode was 10-fold cross-validation. The time taken to build the model is 0.07 seconds. The summary was mentioned in Table 2, and detailed accuracy by class was mentioned in Table 3.

Table 2: Summary

Summary	Values	%
Correctly Classified Instances	263	76.23%
Incorrectly Classified Instances	82	23.76%
Kappa statistic	0.3073	
Mean absolute error	0.2991	
Root mean squared error	0.4368	
Total Number of Instances	345	

Table 3: Detailed Accuracy by Class

Detailed Accuracy By Class									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.386	0.109	0.548	0.386	0.453	0.315	0.666	0.409	No
	0.891	0.614	0.809	0.891	0.848	0.315	0.666	0.81	Yes
Weighted Avg.	0.762	0.485	0.743	0.762	0.747	0.315	0.666	0.708	

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

It was observed from Table 2 that 76.23 % of the total 345 instances were correctly classified instances, and 23.76 % were incorrectly classified instances. Kappa Statistics value was 0.3073, the mean absolute error was 0.2991, and the root mean squared error was 0.4368.

Table 3 described the detailed accuracy by class. For "Yes" class, True Positive Rate (TP Rate) was 0.891, False Positive Rate (FP Rate) was 0.614, Precision was 0.809, Recall was 0.891, F- measure was 0.848, The Matthews correlation coefficient (MCC) was 0.315, ROC Area was 0.666 and Precision-Recall curve (PRC) Area was 0.81. For "No" class, True Positive Rate (TP Rate) was 0.386, False Positive Rate (FP Rate) was 0.109, Precision was 0.548, Recall was 0.386, F- measure was 0.453, The Matthews correlation coefficient (MCC) was 0.315, ROC Area was 0.666 and Precision-Recall curve (PRC) Area was 0.409.

The top 10 rules developed were mentioned in Table 4.

Table 4: Developed rules

Rule No.	Rules	Remarks	Accuracy of Rule
1	TravelComfort = Comfortable AND TravelTiming = TimeSaving AND TravelSafety = Safe: Yes (124.0/9.0)	The commuters who feel comfortable, time-saving, and safe traveling on public transport intend to shift or travel by public transport in the coming days.	93 %
2	TravelConvenience = Convenient AND TravelTiming = TimeSaving AND TravelPerception = Good AND TravelSafety = Neutral AND Purpose = Work AND Educationlevel = Tenth: Yes (16.0/3.0)	The commuters who feel convenient, time-saving, have a good perception of public transport, use public transport for work, and have 10^{th} qualifications have their intention to shift or travel by public transport in the coming days	82%
3	TravelConvenience = Convenient AND TravelTiming = TimeSaving AND TravelPerception = Good: Yes (35.0/5.0)	The commuters who feel convenient, time-saving, and have a good perception of public transport have the intention to travel by public transport in the coming days	86%
4	TravelComfort = Comfortable AND Mode = SharedPT AND TravelTiming = Neutral: Yes (30.0/7.0)	The public transport commuters who feel comfortable traveling in public transport have the intention to travel by public transport in the coming days	77%
5	TravelPerception = Good AND TravelTiming = TimeTaking: Yes (21.0/1.0)	The commuters have a good perception of public transport and time taking and also have the intention to travel by public transport in the coming days.	96%
6	Annualincome = LowIncome AND TravelConvenience = Neutral AND TravelTiming = Neutral AND TravelComfort = Neutral AND TravelPerception = Neutral AND Age	The adult public transport commuters having low income don't want to travel on public transport in the coming days	79%

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Rule No.	Rules	Remarks	Accuracy of Rule
	= Adult AND Mode = SharedPT: No (19.0/4.0)		
7	Annualincome = LowIncome AND Age = Young AND TravelPerception = Neutral: No (7.0/1.0)	The young commuters having low income don't want to travel by public transport in the coming days	86%
8	Annualincome = LowIncome AND TravelSafety = Safe AND Age = Adult AND Mode = PersonalVeh AND TravelPerception = Good: Yes (10.0/2.0)	The adult Personal Vehicle commuters with low income who feel safe and have a good perception having the intention to travel in public transport in the coming days	80%
9	Annualincome = LowIncome AND TravelSafety = Safe AND Mode = SharedPT AND TravelTiming = Neutral AND TravelConvenience = Convenient AND TravelPerception = Good AND Gender = Male: Yes (9.0/2.0)	The public transport male commuters, who have low incomes, feel safe, convenient, and have a good perception, and they want to travel in public transport in the coming days.	78%
10	Annualincome = LowIncome AND TravelSafety = Neutral AND TravelTiming = TimeSaving AND Educationlevel = Graduate AND Gender = Female: Yes (6.0/2.0)	Female commuters with low incomes and graduate qualifications feel that public transport saves time, and they intend to travel in public transport in the coming days.	67%

DISCUSSION

Rule 1 underscores the role of service quality factors—comfort, time efficiency, and safety—in shaping future public transport intentions. These are foundational determinants in mode choice research. Redman et al. (2013) found that service quality is essential in attracting car users to public transport, particularly attributes like comfort and safety. Similarly, Eboli and Mazzulla (2011) emphasized that these subjective experiences strongly influence user satisfaction and continued usage.

Rule 2 integrates both service-based and demographic variables. This includes education level and work-based travel purposes, which reflect findings from Kamargianni and Polydoropoulou (2014), who suggested that students and workers with moderate education are more likely to shift modes when services are convenient and affordable. Positive perception also reflects the growing use of latent variables like attitude and trust in transport modeling (Walker & Ben-Akiva, 2002; Vij & Walker, 2016).

Rule 3 reinforces that perceived convenience and efficiency, along with positive perception, are sufficient motivators for continued use. According to Cao and Cao (2017), perception can outweigh even objective factors like cost or trip duration, especially when trust in the service is established.

Rule 4 points to habitual satisfaction as a strong predictor of continued use. It aligns with the argument by Susilo and Axhausen (2014) that comfort and established routines are powerful motivators for transit loyalty. Moreover, Gärling and Axhausen (2003) showed that habits, once formed under comfortable conditions, resist change.

Rule 5 suggests that positive perception can override the negative impact of time-consuming travel, which is consistent with findings by Cao and Cao (2017). It emphasizes the psychological dimension in travel decision-making—where belief in the system's value matters more than actual performance under some conditions.

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Rule 6 shows that low-income adults may be disillusioned or dissatisfied despite currently using public transport. Such users may feel marginalized or underserved, as noted by Kamargianni and Polydoropoulou (2014). They emphasized that financial hardship alone doesn't ensure loyalty to public transit—quality and dignity of experience also matter.

Rule 7 shows that low-income youth may have aspirational preferences or be discouraged by stigma or inconvenience. According to Jou et al. (2010), younger commuters prioritize speed and comfort, and if public transport fails in those areas, they may prefer alternative modes, even if costlier.

Rule 8 provides optimism—suggesting that even low-income car users are open to switching if they feel safe and hold positive views. This reflects the findings of Redman et al. (2013), who emphasized perceived service quality as a greater lever for behavior change than demographics alone.

Rule 9 aligns with gendered insights in transport research. Men, while less concerned with social stigma than women, still value convenience and safety (Kamargianni & Polydoropoulou, 2014). Positive perception further boosts intention, again affirming the psychological basis for transport loyalty (Vij & Walker, 2016).

Rule 10 uniquely combines gender, education, and efficiency. Educated female commuters likely assess transport rationally, and time efficiency becomes paramount. Gender-sensitive studies (e.g., Gärling & Axhausen, 2003) highlight that efficient travel with minimized risk or delay is a critical factor for women, especially for low-income earners juggling multiple roles.

CONCLUSIONS

The present study of commuter behavior was conducted through decision rules derived from data mining techniques reveals understanding of the factors influencing the intention to shift towards public transport, travel comfort, time savings, safety, travel convenience and positive perception of public transport consistently emerge as important factors to use in coming days.

The factors particularly income, age, education level, and gender, also play a important role in shaping travel intentions. Low-income commuters, both young and adult, show reluctance to continue using public transport unless conditions of safety and convenience are assured. Female and male users demonstrate different priorities, suggesting that gender-sensitive transportation policies are needed. The commuters using private vehicles show a potential to shift to public transport, provided their perceptions are positive, and service quality meets expectations.

The established rules enhance academic knowledge and provide practical guidance for transport planners seeking to improve public transport systems and boost user adoption. A targeted improvement in service quality and campaigns to enhance public perception can significantly influence the modal shift toward sustainable urban mobility.

REFERENCES

[1] Ahmed, S., Rahman, M., & Hasan, S. (2020). Factors affecting public bus ridership in Southeast Asia: A data mining approach. *Journal of Public Transportation, 23*(1), 1–16.

[2] Anwar, S. W., & Bandyopadhyaya, R. (2024). Comparing Public Transport (PT) Travel Perceptions for PT and Personalized Vehicle Users in Well-connected City - Case study of Kolkata, India. *Advances in Transportation Studies*.

[3] Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. MIT Press.

[4] Cao, X., & Cao, J. (2017). The role of perceptions in the acceptance of public transport: A structural equation modeling approach. *Transportation Research Part A: Policy and Practice, 101*, 70–79. https://doi.org/10.1016/j.tra.2017.05.009

[5] Eboli, L., & Mazzulla, G. (2011). A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. *Transport Policy, 18*(1), 172–181.

[6] Gärling, T., & Axhausen, K. W. (2003). Introduction: Habitual travel choice. *Transportation, 30*, 1–11.

[7] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations, 11*(1), 10–18.
[8] Jain, S., Tiwari, G., & Rao, K. V. K. (2013). Modeling commuters' preferences for public transit using decision tree approach. *Procedia - Social and Behavioral Sciences, 104*, 553–562.
[9] Jou, R.-C., Hensher, D. A., Liu, Y.-C., & Chen, C.-C. (2010). Understanding the intention of public transit riders to continue using public transport: A case study in Taipei metro. *Transportation Research Part A: Policy and Practice, 44*(10), 709–720.
[10] Kamargianni, M., & Polydoropoulou, A. (2014). Travellers' behaviour in innovative transport systems: The case of bike-sharing in Athens, Greece. *Transportation Research Procedia, 4*, 328–341. https://doi.org/10.1016/j.trpro.2014.11.025
[11] Karlaftis, M. G., & Vlahogianni, E. I. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities, and some insights. *Transportation Research Part C: Emerging Technologies, 19*(3), 387–399.
[12] Kumar, P., & Vanajakshi, L. (2015). Application of decision tree for mode choice analysis in a developing country. *Procedia Computer Science, 57*, 1044–1053.
[13] Mohammad, F., Rashid, M., & Hossain, M. S. (2017). User satisfaction classification of public transport: A case study using PART classifier. *International Journal of Transportation Science and Technology, 6*(2), 129–143.
[14] Redman, L., Friman, M., Gärling, T., & Hartig, T. (2013). Quality attributes of public transport that attract car users: A research review. *Transport Policy, 25*, 119–127.
[15] Schwanen, T., Dijst, M., & Dieleman, F. M. (2004). The impact of ICT on the location and mobility choices of individuals and firms. *Environment and Planning B: Planning and Design, 31*(6), 759–776.
[16] Shiftan, Y., Outwater, M. L., & Zhou, Y. (2008). Transit market research using structural equation modeling and attitudinal market segmentation. *Transportation Research Record, 2046*(1), 57–63.
[17] Susilo, Y. O., & Axhausen, K. W. (2014). Stability and recurrence of individual's activity-travel pattern: A study using Herfindahl-Hirschman Index. *Transportation Research Part A: Policy and Practice, 64*, 109–127.
[18] Vij, A., & Walker, J. L. (2016). Integrated choice and latent variable models: Insights from a study of U.S. rail commuters. *Transport Policy, 41*, 147–158.
[19] Walker, J. L., & Ben-Akiva, M. (2002). Generalized random utility model. *Mathematical Social Sciences, 43*(3), 303–343. https://doi.org/10.1016/S0165-4896(02)00023-9
[20] Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). Morgan Kaufmann.
[21] Yap, M. D., Cats, O., & van Arem, B. (2018). Crowding valuation in urban trains: Methodological considerations and preliminary results from Netherlands. *Transportation Research Part A: Policy and Practice, 103*, 64–72.
[22] Zhong, C., Arisona, S. M., Huang, X., Batty, M., & Schmitt, G. (2016). Detecting the dynamics of urban structure through spatial network analysis: The case of London. *Transportation Research Part C: Emerging Technologies, 58*, 12–26.
[23] Zhou, F., Zhang, J., Zhao, Z., & Liu, H. (2019). An urban travel mode classification model based on decision trees and survey data. *Journal of Transport Geography, 79*, 102478.