ADROID Conference Series: Engineering and Technology
Volume 2, Issue 2, pp.31-38, 2026

Smarttour: A Hybrid Rule-Machine Learning Based Framework With an Adaptive Refinement Loop For Tourist Behaviour Analysis and Classification

Priya, Shivam Sharma

Department of Computing Technologies, SRM Institute of Science and Technology, Chennai, India

pp3207@srmist.edu.in, ss9993@srmist.edu.in

Abstract. SmartTour is a modular and scalable framework aimed at analyzing and classifying travel behaviour patterns for tourists and population using data relating to mobility. This includes traveling by different modes such as pedestrians, vehicles, etc. The data collected by the SmartTour framework is used for feature extraction concerning travel behaviour, which is then classified using rules and other intelligent techniques such as machine learning. This also includes the extraction of meaningful features such as zone diversity, weekend travel behaviour, temporal entropy, etc. For improving robustness and generalization, SmartTour includes a feedback-based refinement technique that detects discrepancies between rule-based classifiers and machine learning approaches. These inconsistencies contribute towards enhanced classification rules. The system effectively facilitates analytical information such as peak travel time, most popular routes, and tourist-dense regions with the help of interactive visualizations. Therefore, SmartTour proves to be an interpretable, adaptive, and extensible system for urban mobility analysis with its potential applications for smart cities and tourism management.

Keywords: urban mobility, tourist classification, behaviour analytics, smart cities, rule-based learning, machine learning

1. INTRODUCTION

Urban mobility data is an important asset in understanding transportation demand, congestion, tourist flows, and infrastructure usage of smart cities [1], [6], [20]. Compared to residents, the mobility of tourists is more varied in terms of geographical distribution, temporal variability, and usage of pay modes of transportation [7], [20]. Estimating tourist data from mobility information helps in enhancing transport infrastructure and usage to support tourists [2], [10], [14].

However, real-world mobility data is normally not available for use because of privacy issues, commercial sensitivity, or ethical considerations [2], [8]. Even so, such data is normally not labeled for tourists or residents to learn from in order for supervised learning to occur [5], [9]. SmartTour can thus serve as a remedy to this problem. SmartTour gives a privacy-preserving and reproducible frame work to model real mobility data, infer valuable behaviour features, and predict users via a hybrid rule-based and machine learning approach [4], [5], [15]. Unlike traditional heuristics, SmartTour learns and adjusts to discrepancies in the rule based and machine learning models in an adaptive manner via mismatch feedback [16], [17]. The framework can further offer valuable analytics insights to support urban planning and tourist-related tasks [3], [11], [19].

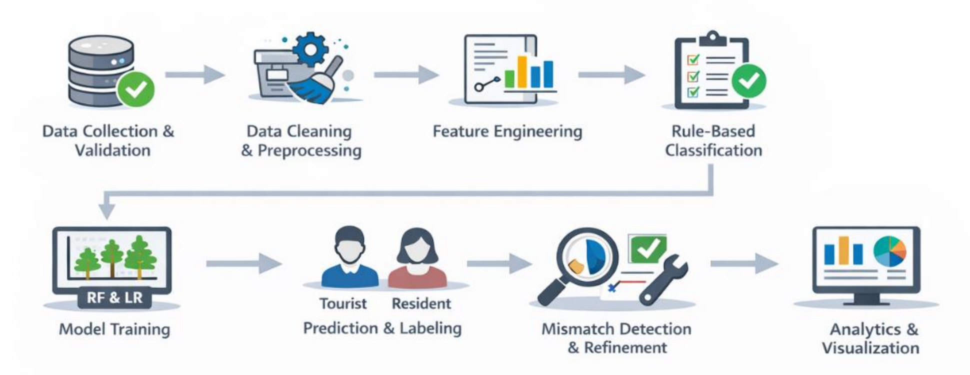

Figure 1. Methodology for Smarttour

2. LITERATURE SURVEY

Recent advancements in urban mobility analytics have increasingly used artificial intelligence and deep learning

techniques to model complex travel behaviour patterns [1], [5]. Graph-based and attention-driven architectures have proven effective in capturing the spatial and temporal dependencies of mobility in urban environments [1], [6]. Transformer-based and clustering-integrated frameworks also improve multi-step mobility prediction and pattern recognition in dynamic trans port systems [4], [5]. These models can learn non-linear interactions among mobility features and adjust to changes over time in urban areas [4], [15]. AI-driven mobility management systems have been explored to optimize urban transportation networks and enhance operational efficiency [2], [19], [20]. These studies emphasize the importance of integrating predictive analytics with traffic control and transportation optimization methods [2], [19]. Generative modeling and forecasting approaches have been introduced to simulate future mobility demand and behavioural trends, helping with proactive planning in smart cities [15], [17]. Agent-based simulations and intelligent modeling techniques further aid in understanding complex mobility interactions in diverse urban layouts [16], [17]. In the tourism field, significant research has focused on analyzing tourist mobility patterns using large-scale datasets and knowledge-driven frameworks [7], [8]. Studies using mobile phone data and extensive movement traces highlight different spatial exploration behaviours among tourists compared to residents [7].
Knowledge graph-based recommendation systems and machine learning approaches have been used to model tourist attraction preferences and enhance personalized travel recommendations [3], [12]. Systematic reviews of artificial intelligence applications in tourism point to the growing integration of predictive analytics, behaviour modeling, and decision-support tools within tourism management systems [8]. Mode choice behaviour and sustainable mobility in urban tourism have also been widely studied [10], [13]. Research on transport mode selection using machine learning methods shows the importance of behavioural features and contextual variables in predicting travel decisions [9], [12]. Studies on sustainable mobility in European capital cities emphasize the impact of urban design, policy changes, and transport accessibility on tourist mobility behaviour [10], [11]. GIS based route optimization and pedestrian network modeling further support sustainable tourism development through data driven planning approaches [14]. Additionally, network-based and smart card data analytics have been used to reveal structural patterns in urban mobility systems [18]. These approaches provide insights into congestion hotspots, travel frequency distributions, and spatial interaction networks, strengthening urban transport management strategies [18], [20]. Overall, the existing literature shows strong progress in mobility prediction, tourism behaviour modeling, AI-based transport management, and sustainable urban planning [1], [2], [7]. However, most approaches either rely solely on complex machine learning architectures with limited clarity or use static rules that lack flexibility [4], [15]. A research gap remains in developing a unified framework that combines clear rule-based reasoning with adaptable machine learning refinement for classifying tourist behaviour in urban mobility systems [16], [17].

3. EXISTING SYSTEM

The various frameworks of analyzing tourist behaviour that currently exist can be categorized based on the following approaches [7], [8].

A. Rule-Based Heuristic

Conventional techniques of identifying tourists use manually evolved heuristics, which rely on domain knowledge [12], [13]. In conventional techniques, the users are classified using thresholds such as the number of days stayed, the number of zones visited, weekend travel, or visit to any known hotspot [7], [10]. However, rule-based systems, even though transparent and easy to interpret, also face many disadvantages, such as: Firstly, the rule-based system is static and does not change with varying travel behaviour patterns [4]. Secondly, threshold settings are often arbitrary and can lead to a biased classification [13]. Thirdly, rule-based systems cannot handle various factors of travel behaviour that are complex [5],[15].

B. Machine Learning-Based Classification Systems

Machine learning techniques typically rely on the application of supervised or unsupervised classifiers to identify tourists based on the extracted mobility features [4], [5]. Supervised models require the availability of datasets with labels. However, such datasets are generally unavailable [9]. Unsupervised classifiers, such as clustering models, struggle to illustrate the semantics and interpretation [4]. Although it can learn non-linear relationships in features, it remains a black box because it is difficult to explain the classification results it produces [5], [15]. Also, once a model has been trained, it does not adapt without being re-trained on new data, which makes it less robust in an urban setting that is constantly changing [1], [6].

C. Limitations of Existing Systems

Furthermore, both rule-based and learning-based systems are standalone systems. There is a lack of flexibility in rule based systems, while there is a lack of transparency and domain control in learning-based systems [4],[5]. Neither system provides a means to track misclassifications over time [15]. Consequently, it shows that currently available systems are not appropriate for long-term deployment in smart cities [2], [19].

4. PROPOSED SYSTEM

SmartTour has introduced a hybrid, adaptive, and explainable framework for analyzing tourist behaviours that transcend the limitations of traditional systems with the integration of reasoning, machine learning, and refinement.

A. Design Philosophy

The designed system is conceived on three major premises:
Interpretability– The decisions should be interpretable and traceable.
Adaptability– The system should grow as the mobility changes.
Scalability– The architecture has to accommodate urban scale .
SmartTour does not change rules in favour of machine learning, it employs machine learning in support of rule-based reasoning [4].

B. Data Processing and Feature Engineering Layer

Generally, raw transport data is trip-based, with fields like time stamps, origin/destination zones, distance traveled, duration, and transport modes [11]. SmartTour engages in data cleaning to remove inconsistencies and invalid entries. From this clean data, the system extracts the following user level behavioural features:

- Number of average trips per day
- Unique zones visited
- Spatial diversity ratio
- Temporal entropy
- Average distance of the trip
- Weekend travel ratio
- Transport mode distribution
- Tourist hotspot visitation ratio

These features together form the behavioural signature for each user [7], [10].

C. Rule-Based Initial Classification

SmartTour applies a set of domain-informed rules in classifying users as either tourists or residents, without utilizing identity information [12]. For example, those users with high spatial diversity and irregular travel patterns are more likely to be tourists [7]. Unlike traditional heuristic systems, SmartTour calibrates rule thresholds based on statistical distributions of the features so that the initial labeling is balanced and unbiased [13]. This serves as a more structured starting point for learning with less dependence on manual labeling [4].

D. Training of the Machine Learning Model

These rule-based classifications form the weak supervision labels for training the machine learning models, including Logistic Regression and Random Forest [5]. Logistic Regression provides interpretability and coefficient-level explanations, while Random Forest captures complex feature interactions and nonlinear behaviour [5]. The models learn latent patterns, not directly encoded by the rules, which improves generalization and robustness [1].

E. Mismatch Detection and Adaptive Rule Refinement

An innovative feature of the SmartTour framework is the “Adaptive Rule Refinement Loop”. Once the training process is complete for the models, the predictions according to the machine learning models and the rule-based models are compared. The areas where the predictions do not match are termed “mismatches”. This information is analyzed to determine if the thresholds set in the rules are too rigid or too flexible. Following this, the parameters of the rules are adjusted automatically [15]. Once this is done, the labels are updated using the improved rules, and the models are retrained [4].

F. Analytics and Visualization Layer

The multiple analytics results are produced by SmartTour, which are:

- Tourist vs resident distribution
- Peak Travel Hours
- Popular tourist routes
- Zone-level Congestion Patterns
- Mode usage trend

Such insights have been depicted using various types of graphs, which makes the system feasible for decision support in urban planning or tourist-related activities [2], [19].

5. METHODOLOGY

The methodology that guides the SmartTour project is built to be a precise, modular, and iterative analytical pipeline that takes raw urban transport data and turns it into useful tourist mobility knowledge.

A. Data Ingestion and Preprocessing

The process starts with the ingestion of urban transport data that has been used, which is supplied in the form of CSV. Each data point consists of information about a trip taken, which includes fields for the time taken, the originating location, the destination, the distance, duration, etc. Preprocessing is an essential operation to ensure the reliability of the data. Missing values of certain attributes, incorrect timestamps, inconsistencies in zone identifiers, and outlier values of travel distances are detected and treated properly. Duplicate values are eliminated to avoid skewed analytics. The time-based characteristics of the hour of the day, the day of the week, and weekend indicators are calculated from the timestamps. This stage ensures that the dataset is clean, standardized, and appropriate for feature extraction and modeling.

B. Feature engineering and Behavioural Analysis

Raw trip-level data does not directly show user behaviour. SmartTour aggregates the data at the user level and primarily focuses on the creation of user behavioural profiles. These features collectively characterize how extensively, how often, and how erratically a user travels, and these factors provide a strong ground for differentiation between residents and tourists. Normalization of features is used for a balanced learning of the model.

C. Rule-Based Initial Classification

The application of the domain-informed rule-based classification layer is the initial decision-making mechanism of SmartTour. These classification rules are devised using existing knowledge of characteristics of tourists' behaviour, which include increased spatial exploration, irregular travel schedules, and inter-zone movements. However, instead of definite thresholds, the parameters in the rules are determined based on statistics, ensuring robustness to variations in datasets. For instance, a group of users above a certain percentile threshold in spatial diversity can be identified as tourists. This step creates interpretable initial labels to use in weak supervision of the machine learning approach, such that the machine learning approach does not rely on any manual labeling or any heuristic mechanism.

D. Machine learning model training

The rule-based labels created during the initial classification phase serve as guidance to train supervised machine learning models, specifically Logistic Regression and Random Forest. This method allows the system to use structured knowledge while also learning deeper patterns from the data. Before training the models, the extracted behavioural features are standardized. This step ensures uniform scaling and prevents features with larger numerical ranges from dominating. The dataset is divided into training and validation subsets, which helps evaluate how well the model can generalize and avoid overfitting. The rule-based classifications act as weak super vision labels. This lets the models learn decision boundaries without needing manually labeled datasets.

Logistic Regression is used because it is easy to interpret. It sets a linear decision boundary in the feature space and provides coefficient estimates for each feature. These estimates show the direction and size of each feature's impact on tourist classification. This makes it easier to understand how behavioural traits like spatial diversity, temporal entropy, or hotspot visits affect predictions. Random Forest is added to work alongside Logistic Regression by modeling nonlinear relationships and complex interactions between features. As an ensemble learning method that uses multiple decision trees, Random Forest lowers variance through bootstrap aggregation and enhances classification strength. It identifies intricate behavioural connections that might not be linear, improving performance in diverse urban mobility patterns. During training, hyperparameters such as the number of trees for Random Forest, maximum tree depth, and regularization strength for Logistic Regression are adjusted to achieve the best results.

Cross-validation techniques are used to ensure stability across different subsets of the data. The trained models uncover hidden behavioural structures that are not explicitly included in the rule-based system. This allows the model to generalize beyond set thresholds and respond to subtle changes in travel patterns. Model performance is measured using various metrics like accuracy, precision, recall, and F1-score. These metrics provide a thorough assessment of classification quality. Consistency measures are also reviewed to check prediction stability across different time segments of the dataset. Together, these evaluation criteria ensure reliability, strength, and clarity in the classification process within dynamic urban mobility environments.

E. Mismatch Detection and Error Analysis

After training, model predictions are compared with rule based labels. Instances where the two disagree are identified as mismatches. Mismatch analysis focuses on understanding:

Which features contributed to misclassification?

Whether rule thresholds are too strict or too lenient.

Whether behavioural patterns have evolved.

This step transforms model errors into valuable feedback, enabling learning beyond static rules or one-time training.

F. Adaptive Rule Refinement Loop

It is the main novelty of SmartTour: the Adaptive Rule Refinement Loop. The thresholds for the rules are automatically updated with statistical feedback from machine learning predictions based on mismatch analysis. Refined rules are reapplied to the dataset for generating updated labels, followed by retraining of models. This iteration may continue until stability in classification is achieved. Unlike traditional systems, SmartTour self-improves with time and without any human intervention, hence making sure that it is adaptable to the continuously changing patterns of mobility while maintaining interpretability.

G. Final Classification and Mobility Analytics

Once convergence is achieved, the final classification results are used to perform detailed mobility analytics. These insights support data-driven decision-making for urban transport planning and tourism management.

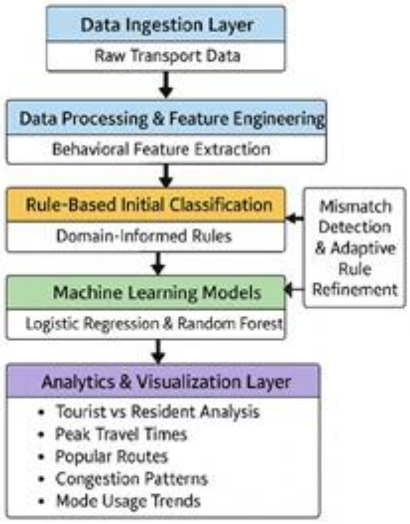

Figure 2. Proposed SmartTour Hybrid Classification Framework

6. SYSTEM ARCHITECTURE

The architectural approach of the SmartTour system is based on a layered modular architecture, which supports the benefits of scalability, interpretability, and extensibility. Its architecture consists of functional layers, which interact with well-defined interfaces. The Data Ingestion Layer involves handling raw data input from transport sources, ensuring it aligns with standard CSV-based data sets. The Data Processing Layer prepares the data by cleaning, normalizing, and extracting features, enabling raw trip-based data to be converted into behavioural representations for users. The Classification Layer involves the application of rule-based logic for the initial classification of tourists and residents. This approach promotes transparent approaches and domain consistency. The Learning Layer involves the application of machine learning models for the identification of complex behavioural relations. This layer is useful for the refinement of the rules for adaptive refinement. Finally, the Analytics and Visualization Layer involves the display of results using different types of charts. Therefore, this architecture ensures fault isolation, reproducibility, and systematic execution of such an analysis pipeline.

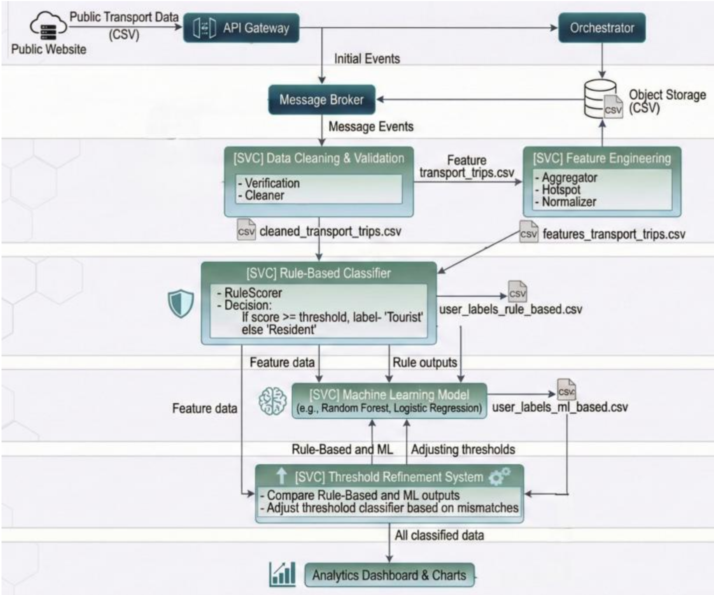

Figure 3. System Architecture

7. PERFORMANCE ANALYTICS AND PROJECT OUTCOMES

The SmartTour system effectively classifies users as tourists or residents based on their urban mobility behaviour. Its hybrid architecture combines rule-based reasoning and supervised machine learning with adaptive refinement. This approach improves classification performance over multiple cycles. Unlike traditional standalone systems, SmartTour enhances its accuracy through continuous feedback between established rules and learned models. Initially, classification starts with rule logic that clearly distinguishes tourists from residents. These rules act as weak supervision for training machine learning models. Logistic Regression and Random Forest models refine the classification boundaries by uncovering hidden behavioural patterns in the mobility data. As the system undergoes successive training and refinement, it shows measurable improvements in classification quality, such as increased accuracy, precision, and recall, along with decreased misclassification rates. A major factor in this improvement is the Adaptive Rule Refinement Loop. In each iteration, the system identifies and examines discrepancies between rule-based predictions and machine learning outputs. These mismatches reveal instances where the rule thresholds might be too strict or not adequately represent changing mobility patterns. The system automatically adjusts the rules based on statistical evidence from the model outputs. This adaptive process increases consistency between classification layers while reducing variance and errors. As a result, the hybrid system proves to be more robust than static heuristic methods or isolated machine learning models. Analyzing the classified data provides important mobility insights. Tourists show greater spatial diversity, visiting multiple areas within the urban network instead of sticking to fixed routes. Their travel times also vary more throughout the day, indicating flexible and exploratory movements. In contrast, residents tend to have structured travel, usually around morning and evening commuting times. The system also differentiates between peak travel periods for tourists and residents. This distinction aids transport scheduling and helps manage congestion. Certain areas within the study network consistently emerge as tourist hotspots because of the high number of users identified as tourists there. Recognizing these zones supports the behavioural logic within the hybrid framework and shows how the approach can be applied in real urban settings.

Beyond improving performance, the SmartTour project creates a fully integrated system for analyzing tourist behaviour. It includes data intake, preprocessing, feature development, classification, adaptive learning, and visualization within a single modular pipeline. One of its main achievements is a privacy-friendly classification method that separates tourists from residents using mobility signatures without needing personal identifiers or demographic data. Another key feature is the explainable rule-based foundation that promotes clear and transparent decision-making. Stake holders can connect classification outcomes to specific behavioural features and thresholds. This clarity is supported by machine learning models that improve predictive accuracy by identifying nonlinear interactions among features and hidden behaviours. The adaptive feedback process ensures that both rule thresholds and learned models keep evolving as mobility patterns change. In addition to analytical results, SmartTour offers visual dashboards and graphics showing mobility trends. This includes the distribution of tourists and residents, peak travel hours, popular routes, congestion patterns by zone, and trends in transport mode usage. The project also provides structured datasets, cleaned mobility records, feature matrices, trained models, and detailed technical documentation. These resources allow for reproducibility, scalability, and easier future system expansion. SmartTour not only fosters methodological innovation with its hybrid adaptive architecture but also presents a workable solution for smart city planning, congestion management, and tourism analysis.

8. CONCLUSION

This paper presents SmartTour, a hybrid and explainable framework for predicting tourist behaviour using urban mobility data. Unlike single-model approaches, the system has a three-tier architecture that combines rule-based reasoning, supervised machine learning, and an adaptive refinement mechanism. This design addresses issues related to unlabeled transport datasets, changing mobility patterns, and the need for clear analytics in smart city applications. The rule-based layer includes seven clear behavioural rules based on features like hotspot visitation ratio, spatial diversity, trip frequency, and travel distance, which allow for traceable classification decisions. To improve predictive power and capture nonlinear relationships, the framework includes Logistic Regression and Random Forest models. A key part of the system is the Adaptive Rule Refinement Loop. This loop compares rule-based predictions with Random Forest outputs to identify mismatches. Among 1,200 users, only eight mismatches were found (five false positives and three missed cases), leading to 99.3% agreement. These discrepancies help adjust thresholds automatically, improving consistency while keeping interpretability. Performance improves steadily at each stage. The standalone rule-based system achieves 94–95% accuracy, which rises to 98–99% after refinement. The Random Forest model alone reaches 97.33% accuracy. The final hybrid ensemble system approaches 99% accuracy and is more robust than single-model methods. Compared to using Random Forest by itself, SmartTour offers better interpretability, adaptability, and less maintenance through threshold tuning instead of complete retraining. By providing insights into tourist-resident distribution, mobility peaks, and hotspot activity, the framework aids sustainable urban transport planning and data-driven smart tourism analytics.

References

[1] Y. Tang, J. He, and Z. Zhao, "Activity-Aware Human Mobility Prediction With Hierarchical Graph Attention Recurrent Network," IEEE Transactions on Intelligent Transportation Systems, vol. 26, no. 2, Feb. 2025.
[2] P. Jahanmanesh, A. Farhadi, and A. Zamanifar, "Mobility Management With AI," IEEE Access, vol. 13, pp. 1–25, 2025, doi: 10.1109/AC CESS.2025.3566337.
[3] J. Xu and Q. Ma, "Tourist Attractions Recommendation by Using Enhanced Location Knowledge Graph," IEEE Access, vol. 13, 2025, doi: 10.1109/ACCESS.2025.3613788.
[4] H. Yang, C. Yan, Z. Chen, and P. Wang, "A K-Shape Clustering Based Transformer-Decoder Model for Predicting Multi-Step Potentials of Urban Mobility Field," IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 8, pp. 10298–10312, Aug. 2024.
[5] M. Chen, Q. Yuan, C. Yang, and Y. Zhang, "Decoding Urban Mobility: Application of Natural Language Processing and Machine Learning to Activity Pattern Recognition, Prediction, and Temporal Transferability Examination," IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 7, pp. 7151–7173, July 2024.
[6] Y. Chen, N. Xie, H. Xu, X. Chen, and D. Lee, "A Multi-Context Aware Human Mobility Prediction Model Based on Motif-Preserving Travel Preference Learning," IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 2, pp. 2139–2152, Feb. 2024.
[7] R. Ke, K. Wang, Q. Liu, L. Liu, and P. Wang, "Uncovering Urban Tourist Mobility Patterns Based on Large-Scale Mobile Phone Data," Journal of Transport Geography, vol. 130, Art. 104496, Jan. 2026.
[8] L. L´opez-Naranjo et al., "Artificial Intelligence in the Tourism Business: A Systematic Review," Frontiers in Artificial Intelligence, vol. 8, Art. 1599391, 2025.
[9] "Machine Learning in Travel Mode Choice Studies: A Systematic Literature Review of Applications, Methods, and Challenges," Results in Engineering, vol. 21, 2025.
[10] L. Zientara et al., "What Drives Tourists' Sustainable Mobility at City Destinations? Insights from Ten European Capital Cities," Journal of Destination Marketing and Management, vol. 33, 2024.
[11] Y.-Z. Liu, H. M. Nguyen, and M. T. Nguyen, "Electric Vehicles and Urban Tourism in Smart Cities: A Bibliometric Review of Sustainable Mobility Trends," World Electric Vehicle Journal, vol. 16, no. 10, 2025.
[12] "Understanding Tourists' Behaviour Toward Transport Mode Choice by Using Machine Learning Methods," European Journal on Decision Processes, vol. 12, 2024.
[13] "A Systematic Review of Mode Choice Behaviour in Urban Transportation," Discover Cities, 2025.

[14] "Promoting Sustainable Urban Tourism Through GIS Route Optimization and Modeling Exclusive Pedestrian Networks," Scientific Reports, 2024.
[15] E. Denteh et al., "Integrating Travel Behaviour Forecasting and Generative Modeling for Predicting Future Urban Mobility," arXiv preprint, 2025.
[16] A. Bahi and A. Ourici, "An Intelligent Agent-Based Simulation of Human Mobility in Extreme Urban Morphologies," arXiv preprint, 2025.
[17] Q. Liu, C. Li, and W. Ma, "GATSim: Urban Mobility Simulation With Generative Agents," arXiv preprint, 2025.
[18] L. Sun, N. Ashrafi, and M. Pishgar, "Optimizing Urban Mobility Through Complex Network Analysis and Big Data From Smart Cards," arXiv preprint, 2025. [19] "Revolutionizing Urban Mobility: AI, IoT, and Predictive Analytics in Adaptive Traffic Control Systems," Electronics, vol. 14, 2025.
[20] "Data-Driven Urban Mobility: Using Machine Learning to Optimize Traffic and Public Transport Operations," Research Study, 2024