

Prediction of Bike Sharing Demand Considering

Built Environment & Natural Factors:

A Case Study of Shenzhen City

物的環境と自然的要因を考慮した

自転車シェアリング需要の予測：

深圳におけるケーススタディ

Department of Urban Engineering, UTokyo 37-236959 WANG DEJIANG

This research focuses on the dockless bike sharing system in Shenzhen, using 7 integrated datasets, including trip records, weather, temporal variances, and built environment data. The OLS model is employed as the baseline model, while XGBoost, Random Forest, DNN, and LSTM are further trained and ensembled into an aggregated model weighted by each model's R-squared performance. To enhance interpretability, SHAP analysis is applied to quantify the contribution and interaction of key features, centering on hour and people flow. Beyond prediction, this research also introduces two novel indices, the Demand Volatility Index (DVI) and the Peak Demand Index (DPI), which are combined into a unified Demand Coordination Priority Index (DCPI). Using the natural break classification, spatiotemporal maps are created to visualize weekday and weekend usage patterns. Based on these findings, this research reveals the different driving factors deciding the bike sharing demand and concludes the different travel patterns of weekdays and weekends.

1. Introduction

The 21st century is defined by a rapid global shift towards urbanization, which, while driving economic growth, places great pressure on urban infrastructure. By 2050, 68% of the world's population is expected to live in cities, an increase of 2.5 billion people from 2018, with this growth concentrated in Asia and Africa^{1,2)}. The velocity of this urban influx often outpaces infrastructure planning, leading to a foundational bias towards car-dependent development and urban sprawl³⁾.

This situation, characterized by low-density and spatially fragmented land use, systemically increases trip lengths and undermines the efficiency of high-capacity public transport⁴⁾. This creates the “first-mile/last-mile” problem, as the challenge of connecting a traveler's origin or destination to a public transit hub⁵⁾. This accessibility gap is an important barrier to public transport vitalization, as spatial accessibility is a crucial factor in a traveler's willingness and choice to

use public transport⁶⁾. Research confirms that many cities are facing the first-mile and last-mile dilemma⁷⁾.

To address this crisis, innovative shared mobility services have emerged as a key solution. Bike Sharing Systems (BSS), in particular, offer a convenient and low-cost option for trips that are too long to walk but too short to drive, effectively extending the accessibility of public transport and bridging the last-mile gap^{8,9,10)}.

A Bike Sharing System (BSS) provides public access to shared bicycles in an urban environment¹¹⁾. Since its inception in 1965, the BSS has evolved through four distinct generations, incorporating technological advancements to improve management and usability. The fourth generation, characterized by QR code access, GPS tracking, and free-floating systems (FFBS), has seen explosive growth around the world, particularly in China, where users grew from 28 million in 2016 to 287 million in 2020¹²⁾.

To address the vehicle dispatching challenges in BSS operations caused by supply-demand mismatches, this research aims to develop and evaluate a high-precision hybrid model for spatio-temporal demand forecasting. This model will provide operators with data-driven decision support to enhance user satisfaction. The study will also identify key external factors (time, weather, urban context), construct and train multiple machine learning and deep learning models, and then develop a superior aggregate model. Using explainable AI, the study will also identify key demand patterns to suggest practical strategies.

This study is significant as it addresses a critical challenge in sustainable urban mobility, aligning with global priorities to improve transport accessibility and social equity^{13,14}. By analyzing a large-scale dataset from Shenzhen, a hyper-dense and rapidly urbanizing megacity, this research develops an interpretable predictive framework to support evidence-based policy planning. The findings offer practical insights for governments and mobility providers in similar Asian megacities facing complex mobility transitions.

2. Literature Review

Bike sharing systems (BSS) are recognized for addressing the "last-mile" problem by enhancing the connectivity of public transport networks¹⁵. This integration yields not only mobility benefits but also positive economic impacts, such as mitigating the negative effect of distance from subway stations on housing prices¹⁶, and significant environmental gains through reduced carbon emissions^{17,18}. However, the primary operational challenge remains the spatio-temporal imbalance of bicycle distribution, driven by the tidal nature of urban commutes¹⁹. This mismatch between supply and demand leads to service failures, necessitating manual intervention through a process called rebalancing. Rebalancing is a complex and costly combinatorial optimization problem, which is hard and particularly challenging for dockless

systems²⁰. Consequently, accurate demand prediction is crucial for efficient and cost-effective rebalancing operations.

The methodologies for BSS demand prediction have evolved significantly, progressing from classical statistical and time-series models like ARIMA to more sophisticated approaches²¹. While early models struggled to capture complex spatial relationships and external factors²², machine learning algorithms such as K-Nearest Neighbors (KNN), Random Forest (RF), and XGBoost demonstrated superior performance by incorporating non-linear relationships and various exogenous factors²³. More recently, deep learning models have become prominent due to their ability to capture deep spatio-temporal dependencies. Some models, like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), excel at modeling temporal patterns, while Convolutional Neural Networks (CNN) and Graph Neural Networks (GNN) are used for spatial correlations^{24,25}. Hybrid models, such as the CNN-LSTM, have been developed to capture these dependencies simultaneously, leading to higher prediction accuracy²⁶.

Despite their high accuracy, the complexity of advanced machine learning and deep learning models often renders them "black boxes," hindering their practical application in high-stakes operational decision-making due to a lack of transparency and trust^{27,28}. To address this, Explainable AI (xAI) techniques have emerged to enhance model interpretability while maintaining performance²⁹. Among various xAI methods, Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive exPlanations (SHAP) are prominent ones. LIME provides local explanations for individual predictions but can miss non-linear relationships³⁰. SHAP, grounded in cooperative game theory, offers a more robust and unified approach by calculating the contribution of each feature to a prediction³¹. Its proven compatibility with a wide range of models, including XGBoost, RF, and deep neural networks, makes it a powerful tool for

building trust and enabling informed, data-driven decisions in BSS operations^{32,33,34}.

3. Methodology and Calculation

This research was conducted in Shenzhen, a rapidly growing Chinese megacity recognized as a pilot for smart and green transportation. The study integrates seven heterogeneous datasets—including bike travel, weather, population heatmaps, Areas of Interest (AOI), and public transport locations—to create a multidimensional profile of urban mobility. The core dataset comprises around 250 million bike-sharing trip records, mainly from January to August 2021, sourced from the Shenzhen government.

To unify these multi-source datasets for analysis, Shenzhen's administrative area was divided into a grid system of 4,233 cells, each approximately 1.23 km by 1.23 km. All data were spatially aggregated to this grid level. Raw bike travel data are filtered to ensure quality, restricting records including selected time period, geographical areas, plausible travel distance, and excluding data from COVID-19 lockdown zones. Extensive feature engineering was performed to create a unified dataset for modeling. Temporal features (e.g., minute, hour, day of the week, month), dummy features (weekends or not, festival or not) weather data (temperature, humidity, wind, precipitation), and urban environmental factors (number of bus/subway stops, standardized people flow, and AOI diversity grade) were generated and assigned to each grid for every 20-minute time interval.

The modeling process began with an Ordinary Least Squares (OLS) model as a baseline, which performed poorly ($R^2 = 0.29$), confirming the presence of complex non-linear relationships in the data. Subsequently, four advanced models were developed, including two machine learning models (XGBoost, Random Forest) and two deep learning models (Deep Neural Network, Long Short-Term Memory). Due to data sparsity after filtering zero-demand records, the dataset was treated as a cross-section, and a random 80/20 split was used for training and testing. Model

performance was evaluated using Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2).

Table 1: An example of table

Model	MSE	MAE	R-Square
OLS	2368.90	27.07	0.29
XGBoost	1298.50	17.25	0.67
Random Forest	1438.23	17.84	0.64
DNN	1512.83	18.94	0.62
LSTM	1554.72	20.02	0.61
Aggregate Model	1292.33	17.05	0.675

All four models significantly outperformed the OLS baseline, with the tree-based XGBoost model demonstrating the highest accuracy. To leverage the strengths of each, a final aggregated model was constructed using a weighted ensemble method. This integrated model achieved the best overall performance, with an MSE of 1292.33, an MAE of 17.05, and an R^2 of 0.675. This final model was then used for interpretation with the SHAP framework.

4. SHAP Analysis

Based on the result of the models, this research employed the SHAP method to interpret the interactions, identifying the key drivers of bike sharing demands and also the inner relationship between specific independent variables.

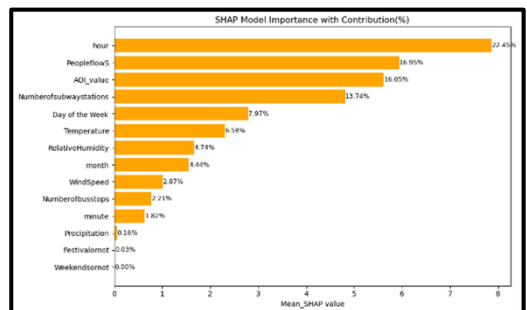


Figure 1: SHAP model global importance

For the global feature importance analysis, Hour is unequivocally the most dominant feature, with its contribution reaching 22.45%, which is far exceeding others, suggesting the decisive role of intraday temporal rhythms in shaping travel demands. Others, including PeopleflowS (standardized people flow, with 16.95% contribution), AOI_value (area function mix index, with 16.05% contribution), Number of subway stations (with 13.74% contribution), constitute the core engine (over 70% contribution in total). However, some features, like precipitation, festival or not, and weekends or not, are far from expectations, with the first two due to the data shortage, and the last one due to being strongly captured by the feature day of the week.

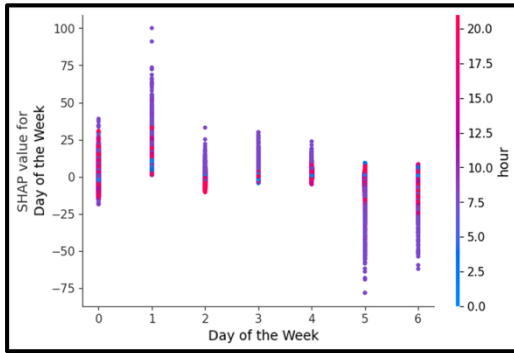


Figure 2: The SHAP dependence plot of day of the week colored by hour

Further interaction analysis revealed the dynamic nature of these feature impacts. Figure 2 demonstrates the strong interaction between Hour and Day of the Week. Weekday (0-4) demand exhibits a typical bimodal commuting pattern, with the strongest positive effect on demand occurring during morning and evening peak hours (red points). In contrast, overall demand is lower on weekends (5-6) (negative SHAP values), with a more subdued, unimodal peak in the afternoon. This plot clearly quantifies the difference between commute-driven and leisure-driven travel patterns.

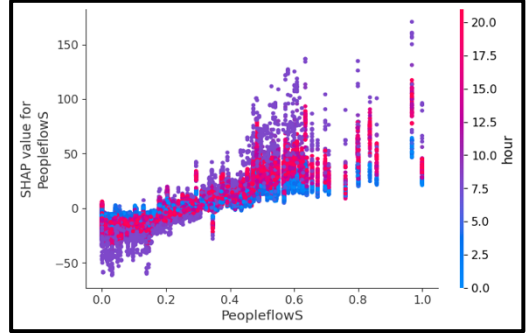


Figure 3: The SHAP dependence plot of people flow colored by hour

Other factors are similarly modulated by time and people flow. This figure 3 mainly focuses on the interaction of peopleflowS (standardized people flow) and hour (variation within one day), confirming that the demand-boosting effect is significantly amplified during peak hours compared to non-peak hours at any given level of people flow. This figure strongly argues against the constant or simple fully people-flow-based redistribution strategy. Instead, a time-weighted redistribution approach could be much better, as resources should be disproportionately allocated to manage high people flow during peak commuting periods because of the highest availability to critical demands in this scenario.

This section mainly does analysis centering the global contribution, the relationship of temporal features with others, and the relationship of people flow with others.

5. Traveling Analysis

To translate raw demand predictions into actionable redistribution strategies, two novel indices were developed: the Demand Volatility Index (DVI), which captures the temporal fluctuation of demand in a grid from the coefficient of variance in statistics, and the Demand Peak Index (DPI), which measures the strength of peak demand relative to the daily average from the peak hour factor in transport engineering. These indices quantify

a grid's stability and peak-hour pressure, respectively. A composite Demand Coordination Priority Index (DCPI) was created by combining DVI and DPI with equal weighting. Using the Natural Breaks (Jenks) method, all grids were classified into five priority levels (from "very stable" to "highly active") based on their DCPI scores. This classification was performed separately for different time slots on weekdays (morning peak, daytime, evening peak, night, and late night) and weekends (morning, high demand period, night, and late night), enabling the development of targeted, time-sensitive redistribution plans based on the different use patterns and inherent characteristics in different time periods.

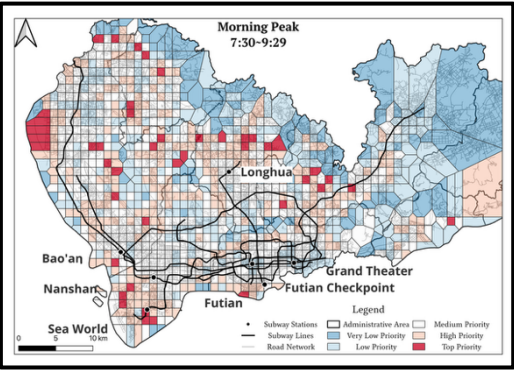


Figure 4: The bike sharing demands on weekdays (Morning Peak)

Figure 4 displays the redistribution priority during the weekday morning peak (7:30-9:29), clearly capturing the city's "tidal" commute characteristics. Highly active areas (in red) are concentrated in core employment hubs like Nanshan High-Tech Park and major residential sub-centers such as Bao'an and Longhua, forming busy transport corridors along subway lines. This visually confirms the critical role of bike-sharing in bridging the "last-mile" gap between homes and metro stations, indicating that redistribution efforts should be deployed along the "residence-metro-workplace" commuter chain.

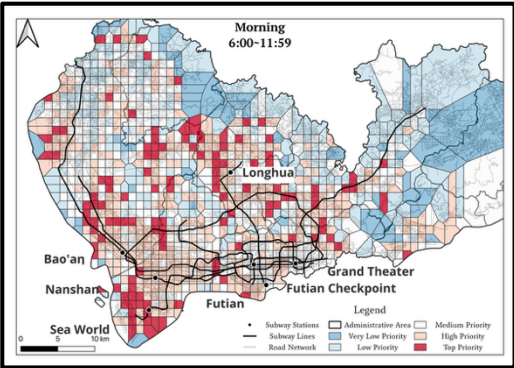


Figure 5: The bike sharing demands on weekends (Morning)

In stark contrast, the weekend morning (Figure 5) reveals a fundamentally different urban rhythm. Instead of a unified tidal flow, demand is decentralized and emerges in scattered, local hotspots. These active areas are no longer dominated by commuter hubs but are instead linked to traditional business districts (Luohu), cross-border checkpoints (Futian), and emerging commercial centers. Trip purposes diversify significantly from commuting to leisure, targeting shopping malls, parks, and cultural venues. This reflects a clear behavioral shift from the individual, efficiency-driven travel of weekdays to group-oriented, experience-driven mobility, where travel destinations are more varied and less predictable.

Based on the analysis of weekday and weekend patterns, there are clear differences in flow dynamics, temporal rhythms, geographical hotspots, and travel motivations. On weekdays, the dominant pattern is a tidal flow between peripheral residential areas and central employment hubs, forming a predictable loop: residents move from suburbs to the core in the morning and return in the evening. In contrast, weekend flows are highly decentralized and event-driven, spreading from multiple residential points toward several popular hotspots and later dispersing back to scattered neighborhoods. Temporally,

weekdays exhibit two sharp and intense peaks corresponding to morning and evening commutes, whereas weekends start later and feature a longer, continuous peak from early afternoon to late evening. Geographically, weekday hotspots are concentrated around major employment centers, such as the Futian CBD and Nanshan High-Tech Park, while during weekends, these same complexes transform into leisure and shopping destinations. In terms of motivation, weekday trips are mainly driven by productive activities (e.g., work), showing high rigidity and time sensitivity, whereas weekend travel is largely driven by consumption-oriented activities (e.g., leisure and shopping) and tends to be more flexible.

5. Conclusion, Discussion, and Future Work

This study employs an explainable ensemble of two machine learning and two deep learning models to reveal the mechanisms behind bike-sharing demand dynamics. Results show that demand and imbalance issues arise from the interaction of temporal, spatial, and activity factors rather than any single driver. Hour is the most influential variable, shaping commuting-driven weekday peaks and flexible leisure-oriented weekend patterns. Subway accessibility and AOI_value strongly affect demand, underscoring the role of bike sharing in last-mile connectivity and the importance of functional diversity over population size. People flow has a direct positive impact, but is amplified by peak periods and mixed-use areas. Temperature effects are non-linear, benefiting demand within optimal ranges. Weekday flows follow predictable tidal commuting patterns, while weekend demands are more scattered around leisure hotspots. These findings highlight the need for integrated rebalancing strategies that consider time, space, and activity simultaneously.

Academically, this study contributes by quantifying the complex synthesis between the built environment and travel behavior, moving beyond linear assumptions to reveal

how the marginal impact of factors like people flow is amplified during peak hours and in functionally diverse areas. Practically, these findings offer actionable insights for urban planners and operators. Key policy recommendations include: prioritizing development within a 500-800m buffer of subway stations to maximize public transport integration, identifying stable, high-demand zones for targeted infrastructure investment, such as dedicated parking, charging stations, and protected bike corridors connecting key hubs, and establishing a government-led platform for information sharing.

The study also acknowledges its limitations, primarily the challenge of inferring causality from correlational data (e.g., the endogeneity between subway stations and demand), data scarcity for certain variables like precipitation, and the single-city focus, which requires further validation for generalizability. Future work should therefore focus on incorporating real-time transport data to capture dynamic substitution effects, exploring causal inference through natural experiments (e.g., using DID or RDD methods to analyze the impact of new metro lines), and extending the analysis to other shared mobility modes to support the development of integrated Mobility-as-a-Service (MaaS) systems.

References

- 1) Farrell, K. and Nijkamp, P. (2019) 'The evolution of national urban systems in China, Nigeria and India', *Journal of Urban Management*, 8(3), pp. 408–419. Available at: <https://doi.org/10.1016/j.jum.2019.03.003>.
- 2) Welford, M.R. and Yarbrough, R.A. (2021) 'Urbanization', in M.R. Welford and R.A. Yarbrough (eds) *Human-Environment Interactions: An Introduction*. Cham: Springer International Publishing, pp. 193–214. Available at: https://doi.org/10.1007/978-3-030-56032-4_8.
- 3) Lesh, M.C. (2013) 'Innovative Concepts in First-Last Mile Connections to Public Transportation', pp. 63–74. Available at: <https://doi.org/10.1061/9780784413210.007>.
- 4) Chetty, V. (2023) 'A Critical Review of Urban Sprawl Studies', *Journal of Geovisualization and Spatial Analysis*, 7(2), p. 28. Available at: <https://doi.org/10.1007/s41651-023-00158-w>.

- 5) Shu, P. et al. (2021) 'Data-driven shuttle service design for sustainable last mile transportation', *Advanced Engineering Informatics*, 49, p. 101344. Available at: <https://doi.org/10.1016/j.aei.2021.101344>.
- 6) Tilahun, N. et al. (2016) 'Transit use and the work commute: Analyzing the role of last mile issues', *Journal of Transport Geography*, 54, pp. 359–368. Available at: <https://doi.org/10.1016/j.jtrangeo.2016.06.021>.
- 7) Kåresdotter, E. et al. (2022) 'First Mile/Last Mile Problems in Smart and Sustainable Cities: A Case Study in Stockholm County', *Journal of Urban Technology*, 29(2), pp. 115–137. Available at: <https://doi.org/10.1080/10630732.2022.2033949>.
- 8) Chen, Y. and Wang, H. (2018) 'Pricing for a Last-Mile Transportation System', *Transportation Research Part B: Methodological*, 107, pp. 57–69. Available at: <https://doi.org/10.1016/j.trb.2017.11.008>.
- 9) Baek, K. et al. (2021) 'Electric scooter sharing: How do people value it as a last-mile transportation mode?', *Transportation Research Part D: Transport and Environment*, 90, p. 102642. Available at: <https://doi.org/10.1016/j.trd.2020.102642>.
- 10) Zuo, T. et al. (2020) 'First-and-last mile solution via bicycling to improving transit accessibility and advancing transportation equity', *Cities*, 99, p. 102614. Available at: <https://doi.org/10.1016/j.cities.2020.102614>.
- 11) Jiang, N., & Cai, J. M. (2017) How Governments Play Their Roles in the Share Economy in the Perspective of Bike-sharing - CNKI. Available at: https://www.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2017&filename=HEAR201704022&uniplatform=OVERSEA&v=e3-9DT1Je0equefpe_plAk51JafUaBaXx4xAld8Qlv2iUMHWrtvkb4HiDuvbgI_X (Accessed: 28 April 2025).
- 12) SONG, S., LIU, A. and MA, J. (2022) STATUS AND OPPORTUNITIES OF SHARED MOBILITY SYSTEMS IN CHINA. Volvo Research and Educational Foundations. Available at: chrome-extension://efaidnbmnnnibpcajpgclefindmkaj/https://vref.se/wp-content/uploads/2022/08/Informal-and-Shared-Mobility-Systems-in-China_16june.pdf.
- 13) OECD (2025) Promoting Active Ageing in Southeast Asia. Available at: https://www.oecd.org/en/publications/promoting-active-ageing-in-southeast-asia_22849f38-en.html (Accessed: 2 June 2025).
- 14) Cheng, L. et al. (2022) 'The Role of Bike Sharing in Promoting Transport Resilience', *Networks and Spatial Economics*, 22(3), pp. 567–585. Available at: <https://doi.org/10.1007/s11067-021-09518-9>.
- 15) Zhu, L. et al. (2022) 'Approaching Sustainable Bike-Sharing Development: A Systematic Review of the Influence of Built Environment Features on Bike-Sharing Ridership', *Sustainability*, 14(10), p. 5795. Available at: <https://doi.org/10.3390/su14105795>.
- 16) Chu, J. et al. (2021) 'The Last Mile Matters: Impact of Dockless Bike Sharing on Subway Housing Price Premium', *Management Science*, 67(1), pp. 297–316. Available at: <https://doi.org/10.1287/mnsc.2019.3550>.
- 17) Ding, M. et al. (2025) 'Environmental Benefits Evaluation of a Bike-Sharing System in the Boston Area: A Longitudinal Study', *Urban Science*, 9(5), p. 159. Available at: <https://doi.org/10.3390/urbansci9050159>.
- 18) Liu, G. et al. (2025) 'Refinement of carbon emission reduction of dockless bike-sharing users' behavior under different scenarios', *Frontiers in Environmental Science*, 13. Available at: <https://doi.org/10.3389/fenvs.2025.1594262>.
- 19) Liang, J., Jena, S.D. and Lodi, A. (2024) 'Dynamic rebalancing optimization for bike-sharing systems: A modeling framework and empirical comparison', *European Journal of Operational Research*, 317(3), pp. 875–889. Available at: <https://doi.org/10.1016/j.ejor.2024.04.037>.
- 20) Bruck, B.P. et al. (2019) 'The Static Bike Sharing Rebalancing Problem with Forbidden Temporary Operations', *Transportation Science*, 53(3), pp. 882–896. Available at: <https://doi.org/10.1287/trsc.2018.0859>.
- 21) Kontopoulou, V.I. et al. (2023) 'A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks', *Future Internet*, 15(8), p. 255. Available at: <https://doi.org/10.3390/fi15080255>.
- 22) Rennie, N. et al. (2023) 'Analysing and visualising bike-sharing demand with outliers', *Discover Data*, 1(1), p. 1. Available at: <https://doi.org/10.1007/s44248-023-00001-z>.
- 23) Gao, C. and Chen, Y. (2022) 'Using Machine Learning Methods to Predict Demand for Bike Sharing', in J.L. Stienmetz, B. Ferrer-Rosell, and D. Massimo (eds) *Information and Communication Technologies in Tourism 2022*. Cham: Springer International

- Publishing, pp. 282–296. Available at: https://doi.org/10.1007/978-3-030-94751-4_25.
- 24) Pan, Y. et al. (2019) 'Predicting bike sharing demand using recurrent neural networks', *Procedia Computer Science*, 147, pp. 562–566. Available at: <https://doi.org/10.1016/j.procs.2019.01.217>.
 - 25) Xiao, G. et al. (2021) 'Demand prediction for a public bike sharing program based on spatio-temporal graph convolutional networks', *Multimedia Tools and Applications*, 80(15), pp. 22907–22925. Available at: <https://doi.org/10.1007/s11042-020-08803-y>.
 - 26) Chamola, V. et al. (2023) 'A Review of Trustworthy and Explainable Artificial Intelligence (XAI)', *IEEE Access*, 11, pp. 78994–79015. Available at: <https://doi.org/10.1109/ACCESS.2023.3294569>.
 - 27) Laqua, A. et al. (2023) 'Exploring User Experience in Sustainable Transport with Explainable AI Methods Applied to E-Bikes', *Applied Sciences*, 13(20), p. 11277. Available at: <https://doi.org/10.3390/app132011277>.
 - 28) Salih, A. et al. (2025) 'A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME', *Advanced Intelligent Systems*, 7(1), p. 2400304. Available at: <https://doi.org/10.1002/aisy.202400304>.
 - 29) Onen, S. et al. (2023) 'Interpretable Machine Learning for Intelligent Transportation in Bike-Sharing', in *2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*. 2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), pp. 1–6. Available at: <https://doi.org/10.1109/ICSTSN57873.2023.10151456>.
 - 30) Lundberg, S.M. and Lee, S.-I. (2017) 'A unified approach to interpreting model predictions', in *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc. (NIPS'17), pp. 4768–4777.
 - 31) Karamanou, A., Kalampokis, E. and Tarabanis, K. (2022) 'Linked Open Government Data to Predict and Explain House Prices: The Case of Scottish Statistics Portal', *Big Data Research*, 30, p. 100355. Available at: <https://doi.org/10.1016/j.bdr.2022.100355>.
 - 32) Wan, M. et al. (2022) 'Application of Random Forest and Shap for Real-Time Taxi Drivers' Traffic Violations Detection and Feature Analysis'. Rochester, NY: Social Science Research Network. Available at: <https://doi.org/10.2139/ssrn.4142196>.
 - 33) Song, H. et al. (2023) 'Elite male table tennis matches diagnosis using SHAP and a hybrid LSTM-BPNN algorithm', *Scientific Reports*, 13(1), p. 11533. Available at: <https://doi.org/10.1038/s41598-023-37746-1>.