aut. Agnieszka Łysak, UJK

18.06.2025

Decoding human behavior from mobility data: AI-driven insights

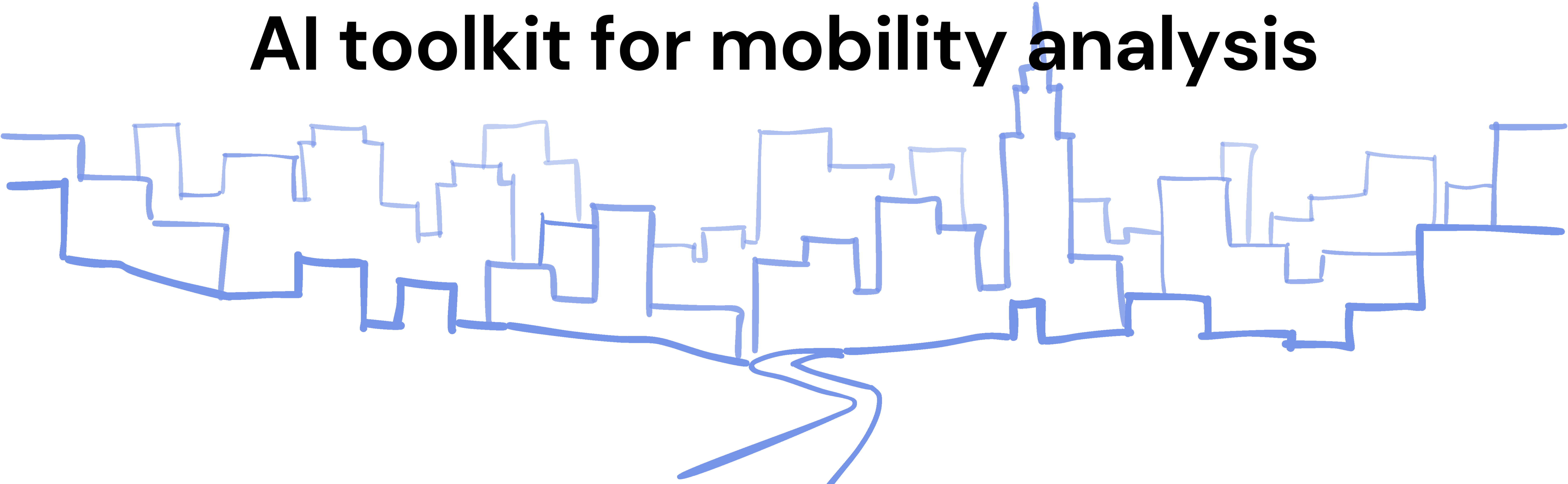
Introduction

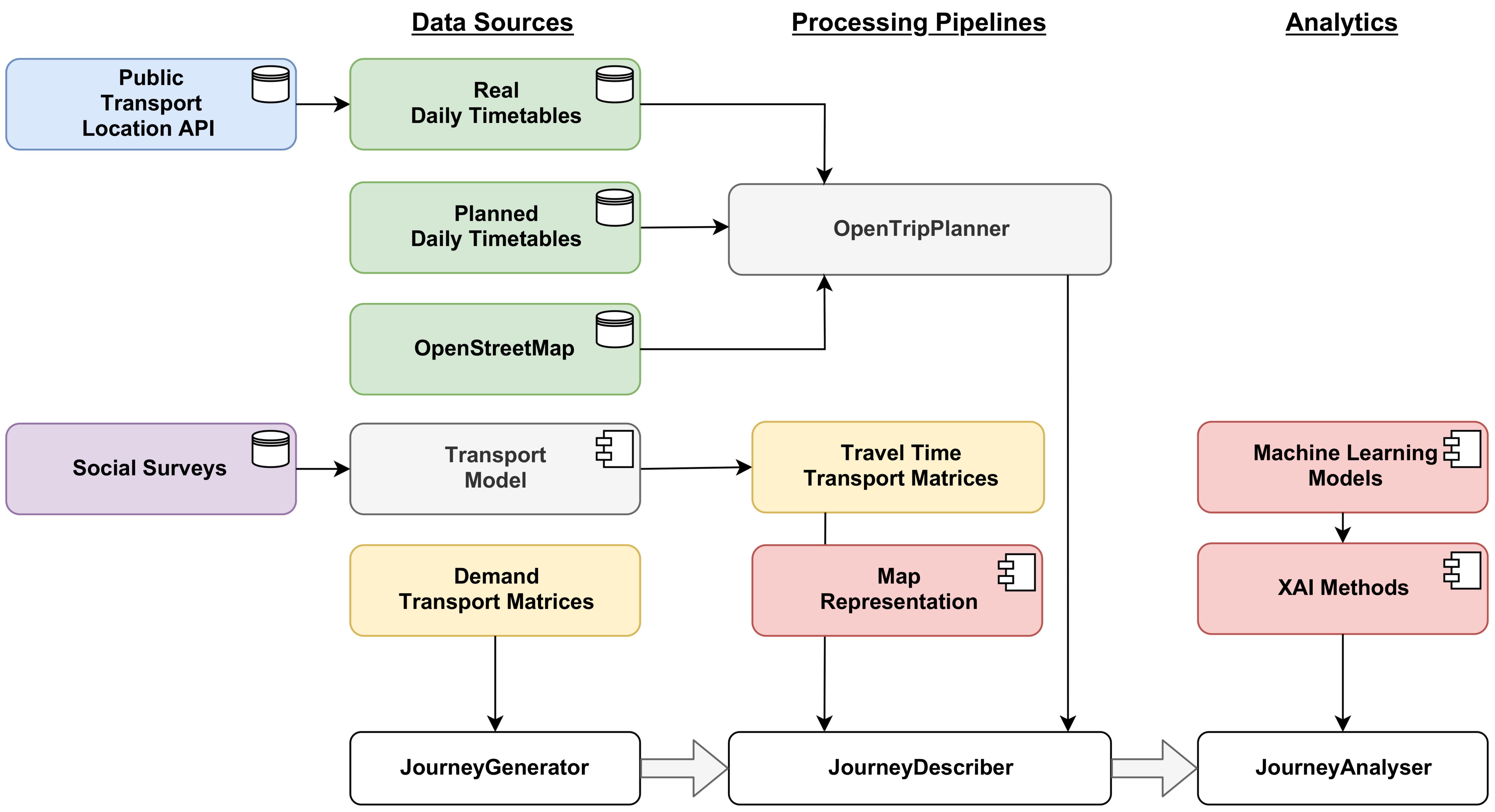
Every day, millions of journeys leave behind a digital footprint – a trail of where we go, when, and how. What can we really learn from these traces?

Can we uncover how we live, work, move? Are there hidden psychological or urban patterns embedded in our mobility?

In this short lecture, I'll introduce you a toolkit I developed as part of my PhD research and demonstrate use cases, where AI and machine learning help decode mobility behavior.
Together, they reveal how data can uncover the habits shaping our cities and daily lives.

AI toolkit for mobility analysis

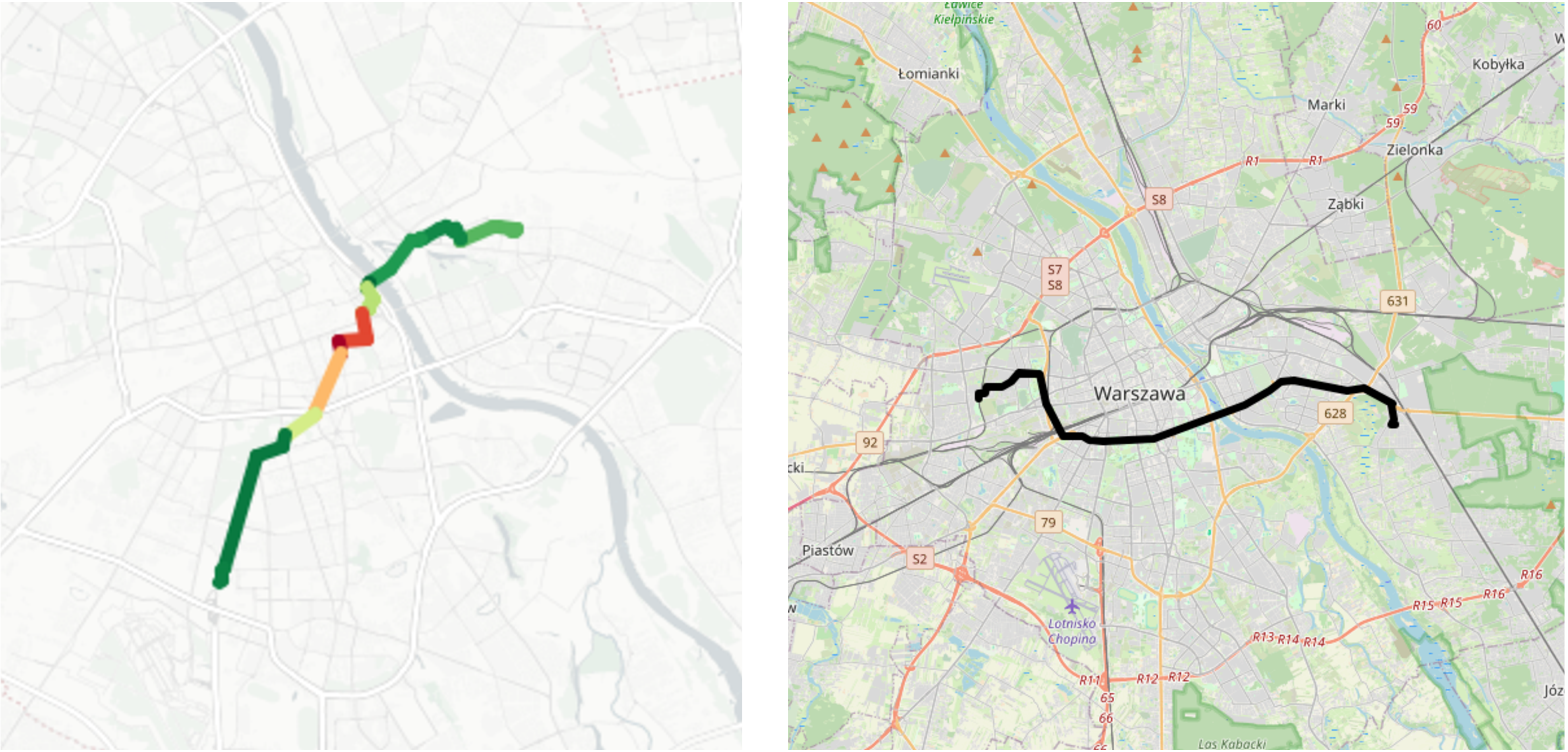




Workflow chart, based on: Luckner, M., Wrona, P., Grzenda, M., & Łysak, A. (2024). Analysing urban transport using synthetic journeys. In Computational Science – ICCS 2024. Springer Nature Switzerland AG, https://link.springer.com/chapter/10.1007/978-3-031-63783-4_10.

id_SURVEY	sex_SURVEY	yearOfBirth_SURVEY	monthOfBirth_SURVEY	education_SURVEY	hasBike_SURVEY	carAvailability_SURVEY	cyclingLimitations_SURVEY	homeAddressLatitude_SURVEY	homeAddressLongitude_SURVEY	startingAtHome_SURVEY
1	FEMALE	1988	SEPTEMBER	BACHELOR	TRUE	TRUE	FALSE	52.3256	21.0597	TRUE
2	MALE	1970	JANUARY	BACHELOR	FALSE	FALSE	FALSE	52.1127	21.0003	FALSE
3	MALE	1951	OCTOBER	MASTER	TRUE	FALSE	TRUE	52.17375	21.0744	TRUE
4	FEMALE	1989	DECEMBER	MASTER	FALSE	TRUE	FALSE	52.6056	21.7514	FALSE
5	MALE	1941	MAY	SECONDARY	TRUE	TRUE	FALSE	52.1561	21.0579	TRUE
6	FEMALE	1948	JULY	SECONDARY	TRUE	TRUE	TRUE	52.1309	21.0579	TRUE
7	MALE	1975	JANUARY	MASTER	FALSE	FALSE	TRUE	52.1544	21.9418	FALSE

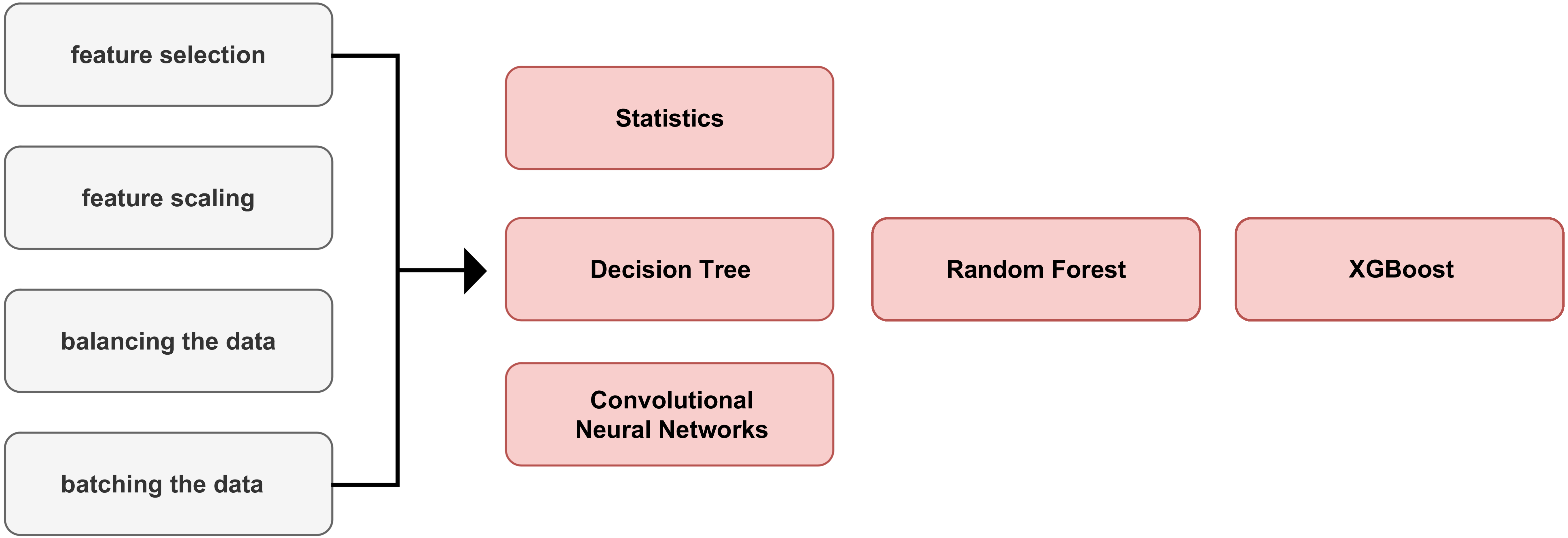
Exemplary tabular representations of the mobility data, author: Marcin Luckner, 2025.



Exemplary map representations of the mobility data, author: Agnieszka Łysak, 2025.

Real data is essential but often limited, incomplete, biased toward certain social group or geographic areas.
That’s why simulations are valuable - the help fill gaps, model interesting events under various conditions and expand datasets, a critical need in machine learning.

Simulations are great tool to test alternative scenarios, but the crucial step is real-world validation, which ensures realism and reliability.



Schema of the machine learning modules, author: Agnieszka Łysak, 2025.

Mobility data is complex, influenced by time, location, transport mode, personal traits, urban structure.

Statistical models provide a strong baseline. They're transparent and effective to more straightforward relationships.

Decision trees work well with mixed data, handles missing values, offering good performance, while remaining interpretable.

Neural networks are powerful for capturing spatial and temporal patterns. They're well-suited for large datasets, though they require explanation and tuning.

Case study 1: the 15-minute city dream

15-minute city is an urban concept based on the idea that all essential amenities should be accessible within a 15-minute reach.[1]

Originally, this proximity was measured by walking distance.

Nowadays the concept expanded to include longer times frames, such as 20-minute city or the broader X-minute city, as well as other sustainable modes of transport like cycling or public transit.

Allam, Z., Nieuwenhuijsen, M., Chabaud, D., & Moreno, C. (2022). The 15-minute city offers a new framework for sustainability, liveability, and health. The Lancet Planetary Health, 6(3), e181–e183.

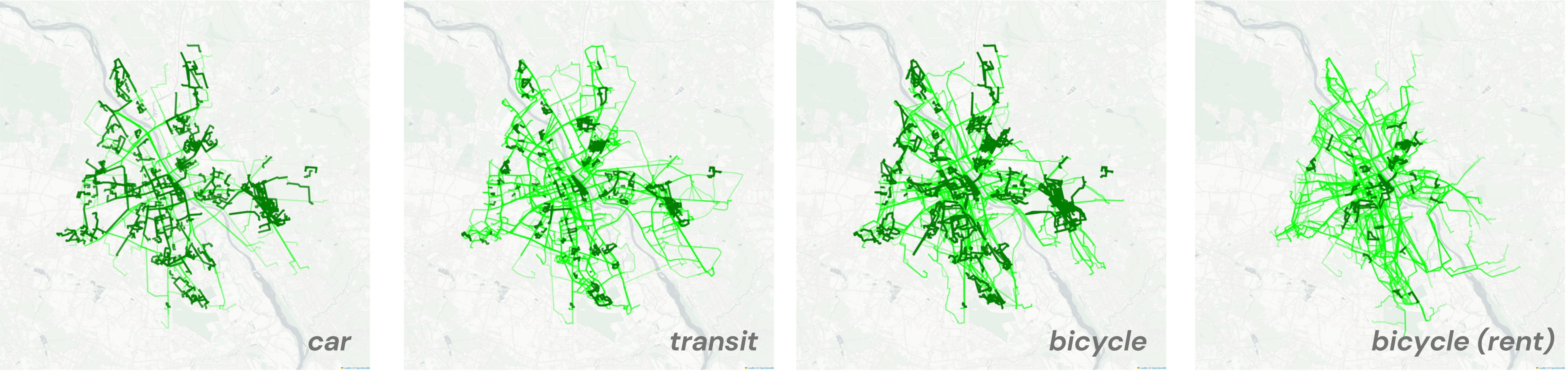
Diagram of trips in the city of Warsaw, black are travels up to 15 minutes, authors: Agnieszka Łysak and Marcin Luckner, 2025.

Our analysis was based on four transportation modes: car, private bicycle, shared bicycle and public transit. The focus was on accessibility within 15-minute time interval.

The AI toolkit was used solely for data generation, after which classical quantitative methods were applied to assess the concept in the city of Warsaw.

Luckner, M., Łysak, A., & Archanowicz-Kudelska, K. (w druku). Modelling 15-minute city work and education amenities using surveys and simulations. W Proceedings of the 32nd International Conference on Information Systems Development (ISD 2024). Gdańsk, Polska, https://aisel.aisnet.org/isd2014/proceedings2024/transformation/22/.

Education amenities:



Work amenities:

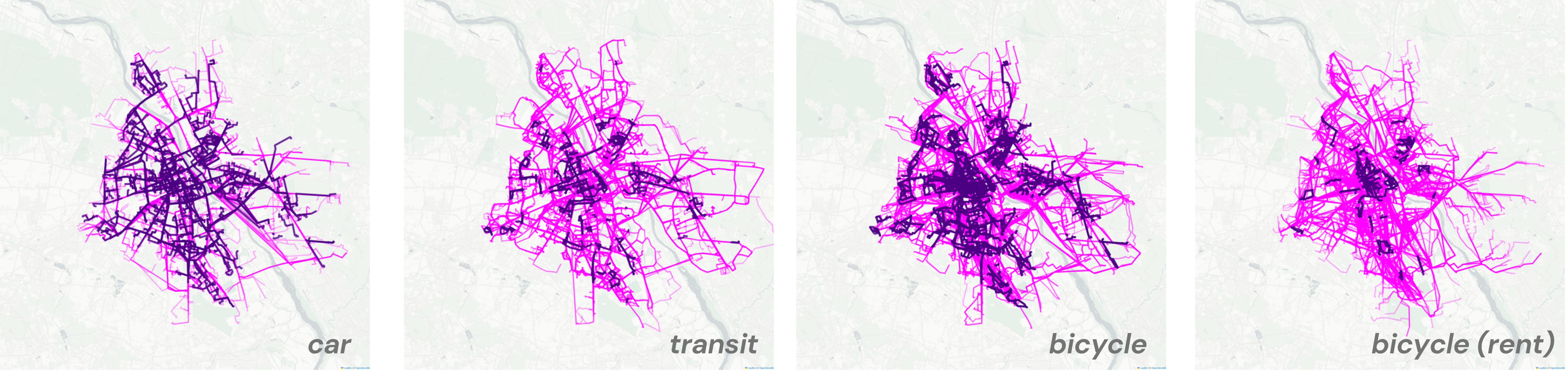
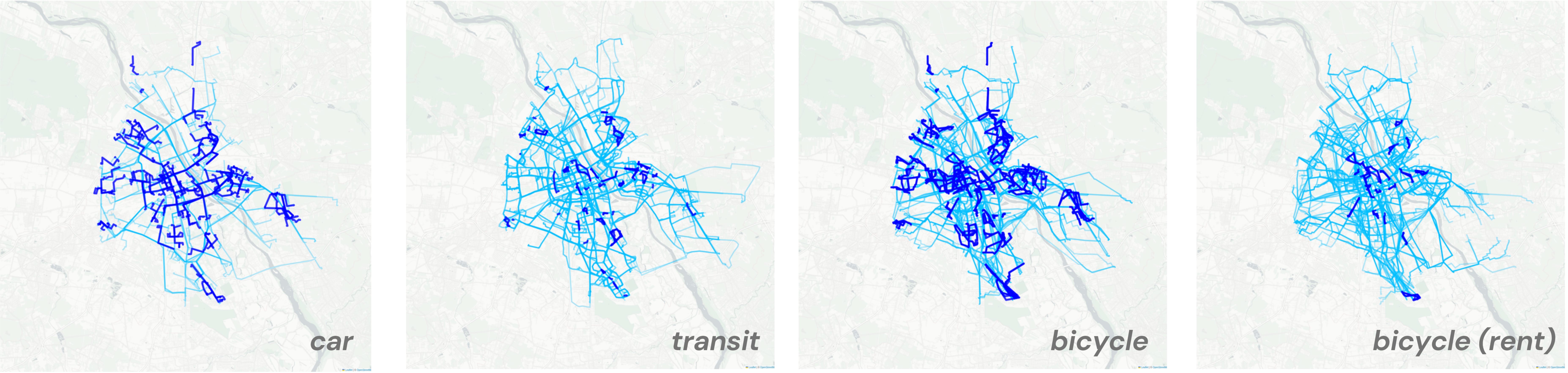


Diagram of trips in the city of Warsaw, travels up to 15 minutes are darker shade, authors: Agnieszka Łysak and Marcin Luckner, 2025.

Healthcare amenities:



Commercial amenities:

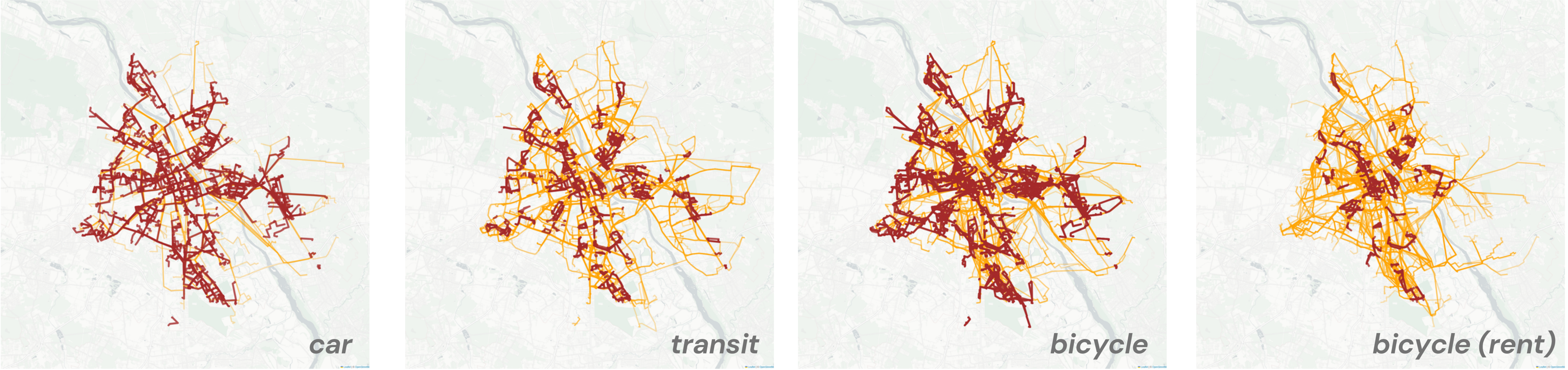
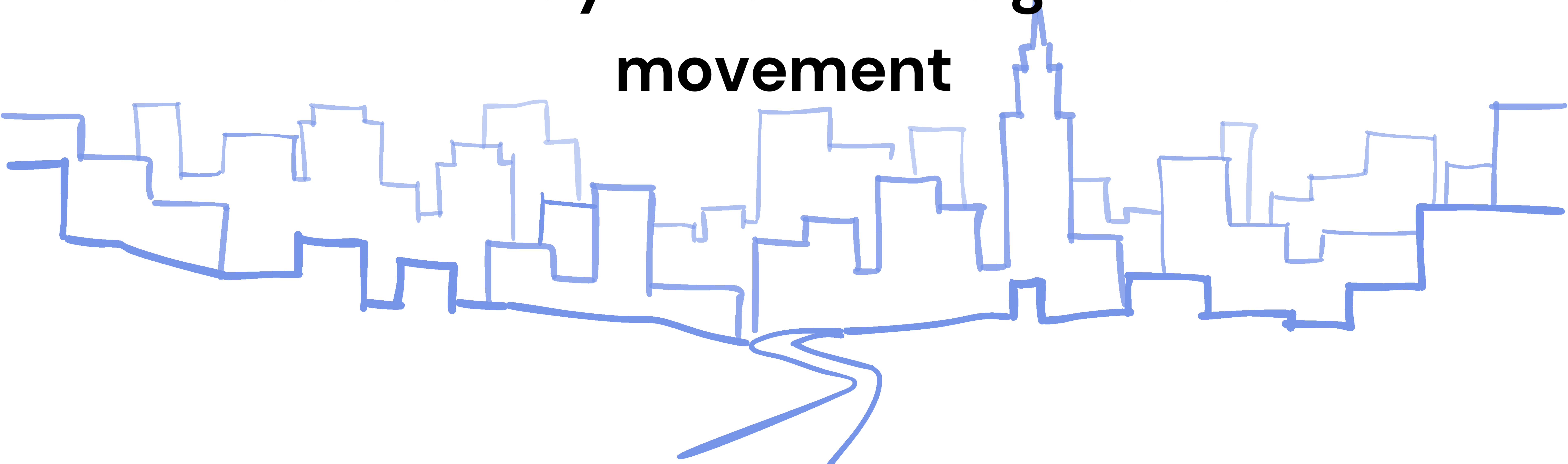


Diagram of trips in the city of Warsaw, travels up to 15 minutes are darker shade, authors: Agnieszka Łysak and Marcin Luckner, 2025.

Case study 2: health insights from movement

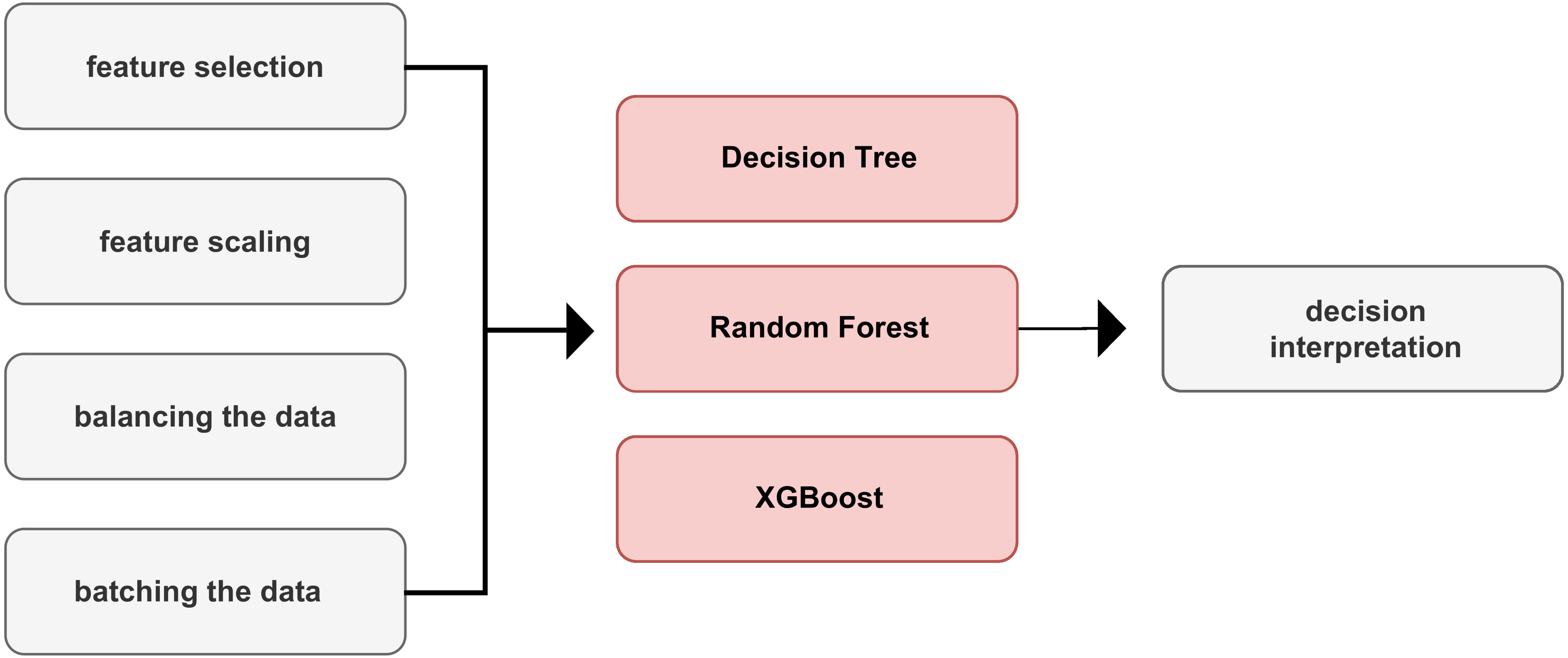


The second application focused on assessing the physical activity of primary school children based on transportation data. Survey responses from parents were isolated, and with the support of simulations, a dataset was constructed.

Machine learning methods were then applied to identify the key factors influencing the choice of active transportation modes.

The AI toolkit was used for data oversampling, followed by the application of machine learning methods based on decision trees.

Luckner, M., Łysak, A., & Archanowicz-Kudelska, K. (w druku). Estimating physical activity during travel to school. W Proceedings of the SoGood 2024: ECML-PKDD Workshop on Data Science for Social Good. Wilno, Litwa.



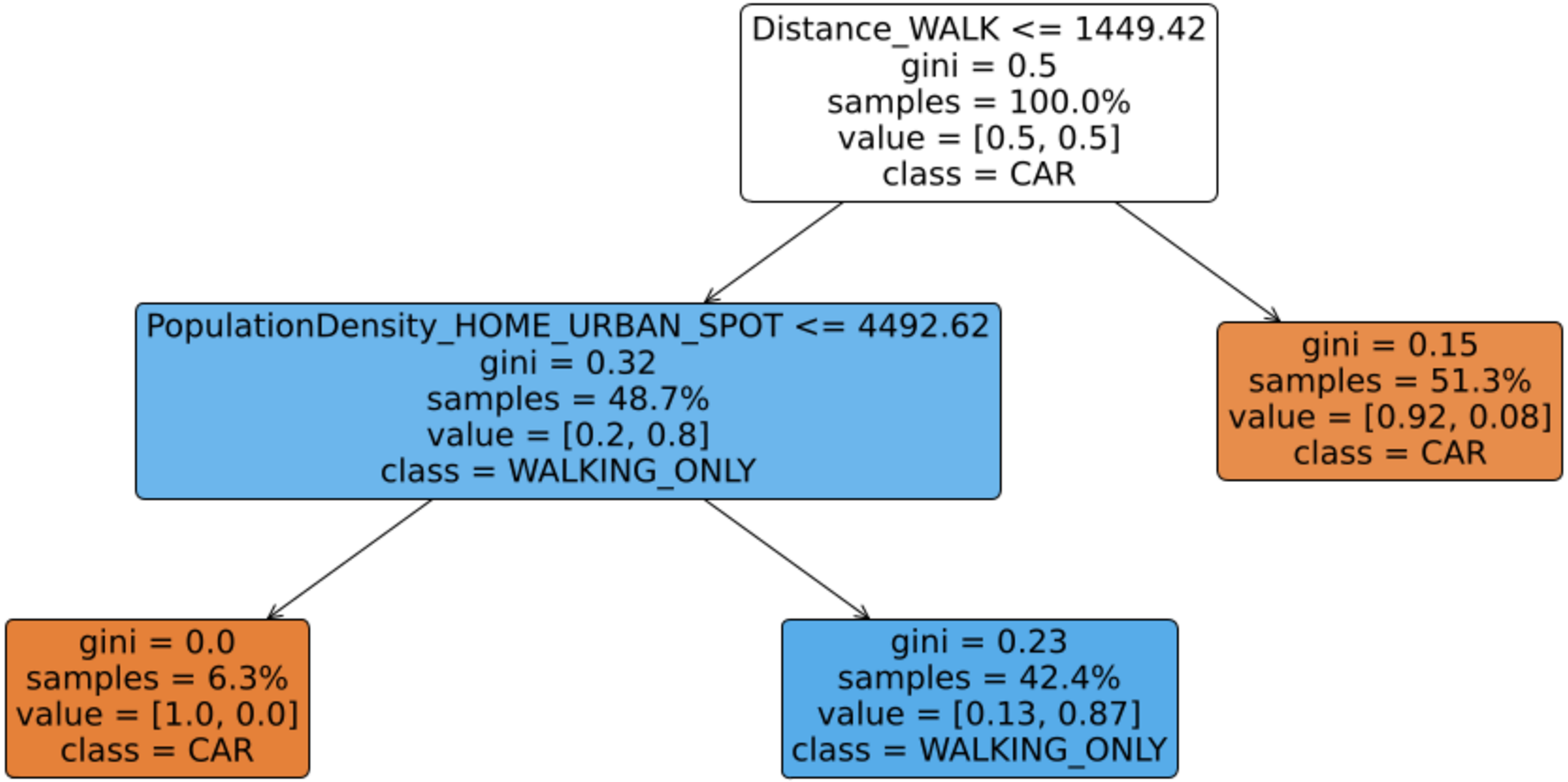
Schema of the machine learning analytics module, author: Agnieszka Łysak, 2024.

Machine learning models were applied to identify features influencing the choice to walk vs. travel by car. The data format was different than previously – only tabular dataset was used.

Analysing school reachability by walking is a key to evaluating children's health, especially as sedentary time and related health issues continue to rise.

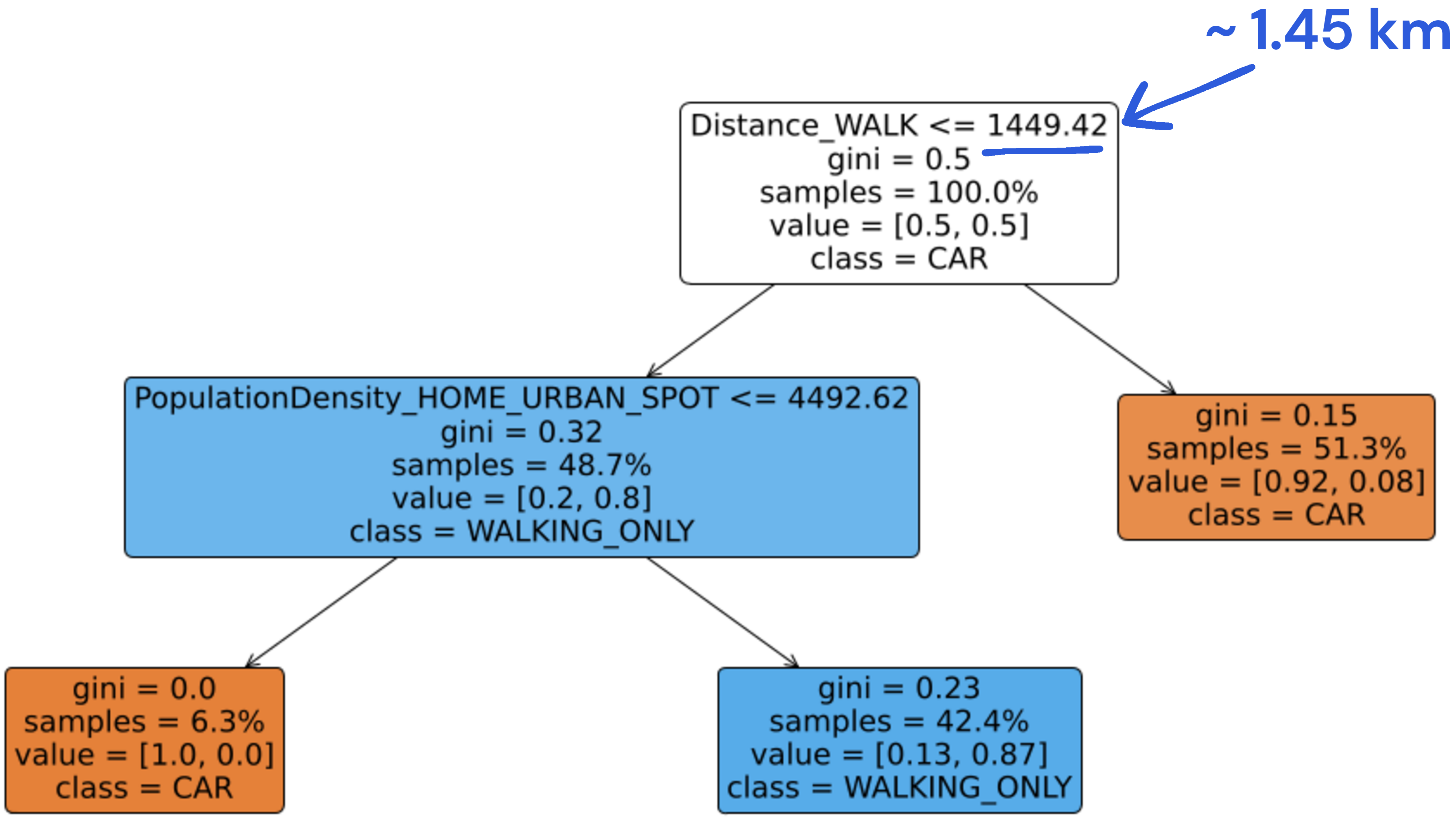
A walk 15-minute walk to school can provide half of the recommended 1 hour of daily physical activity needed for good health.

Luckner, M., Łysak, A., & Archanowicz-Kudelska, K. (w druku). Estimating physical activity during travel to school. W Proceedings of the SoGood 2024: ECML-PKDD Workshop on Data Science for Social Good. Wilno, Litwa.



The decision tree graph, showing how car or walk transportation mode is being classified.

Luckner, M., Łysak, A., & Archanowicz-Kudelska, K. (w druku). Estimating physical activity during travel to school. W Proceedings of the SoGood 2024: ECML-PKDD Workshop on Data Science for Social Good. Wilno, Litwa.



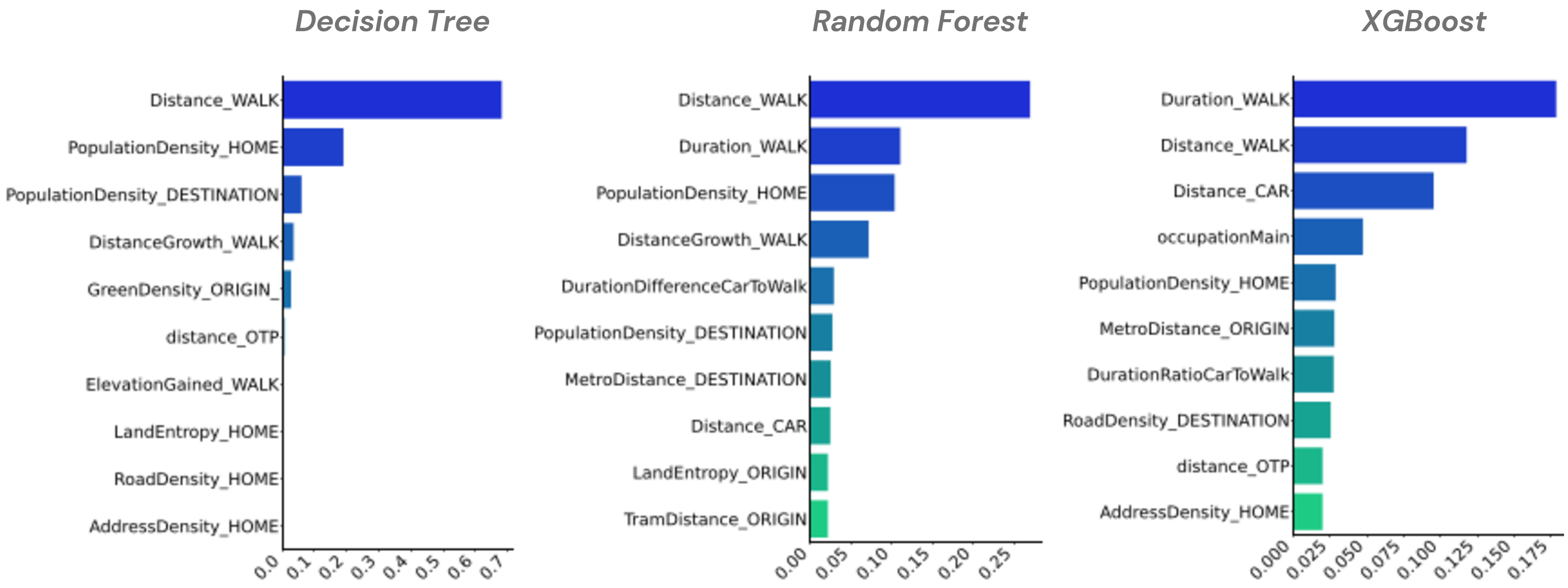
The decision tree graph, showing how car or walk transportation mode is being classified.

Luckner, M., Łysak, A., & Archanowicz-Kudelska, K. (w druku). Estimating physical activity during travel to school. W Proceedings of the SoGood 2024: ECML-PKDD Workshop on Data Science for Social Good. Wilno, Litwa.

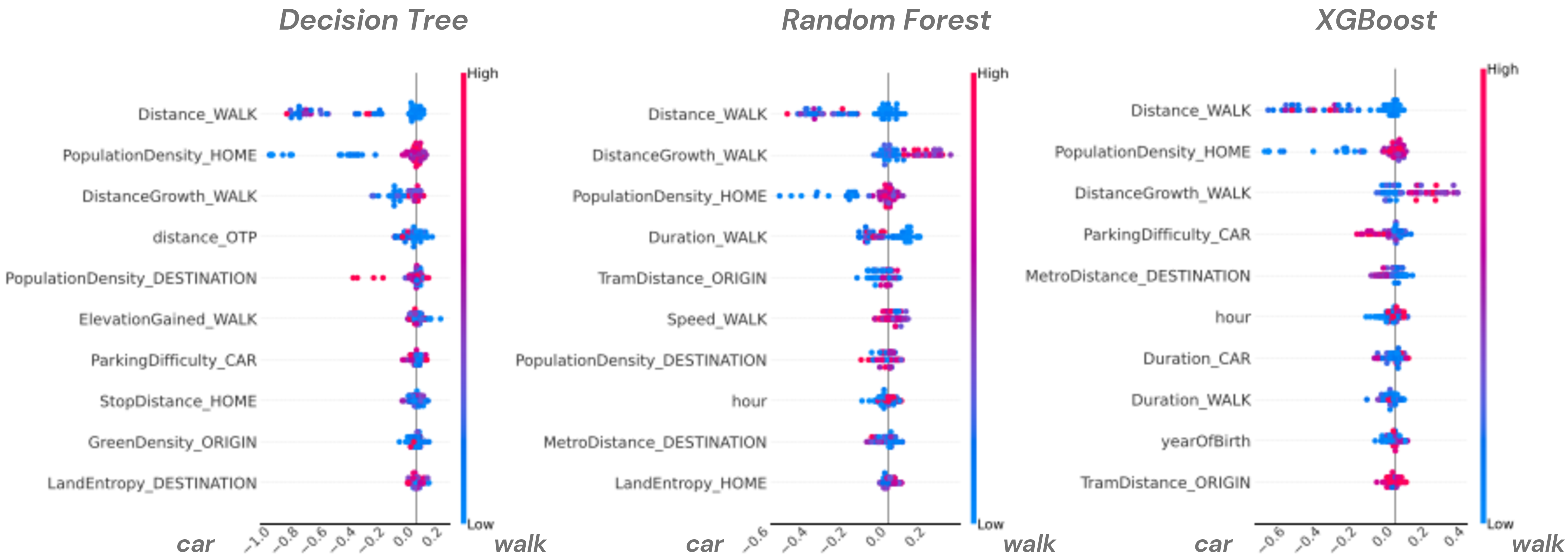
The decision tree analysis showed that distance of up to 1.45 km to school is acceptable when the neighborhood is densely populated. In that case walking to school was choosen.

Otherwise, in less populated areas or when the walking distance exceeds the threshold, parents travelling with children tend to use a private car.

Luckner, M., Łysak, A., & Archanowicz-Kudelska, K. (w druku). Estimating physical activity during travel to school. W Proceedings of the SoGood 2024: ECML-PKDD Workshop on Data Science for Social Good. Wilno, Litwa.

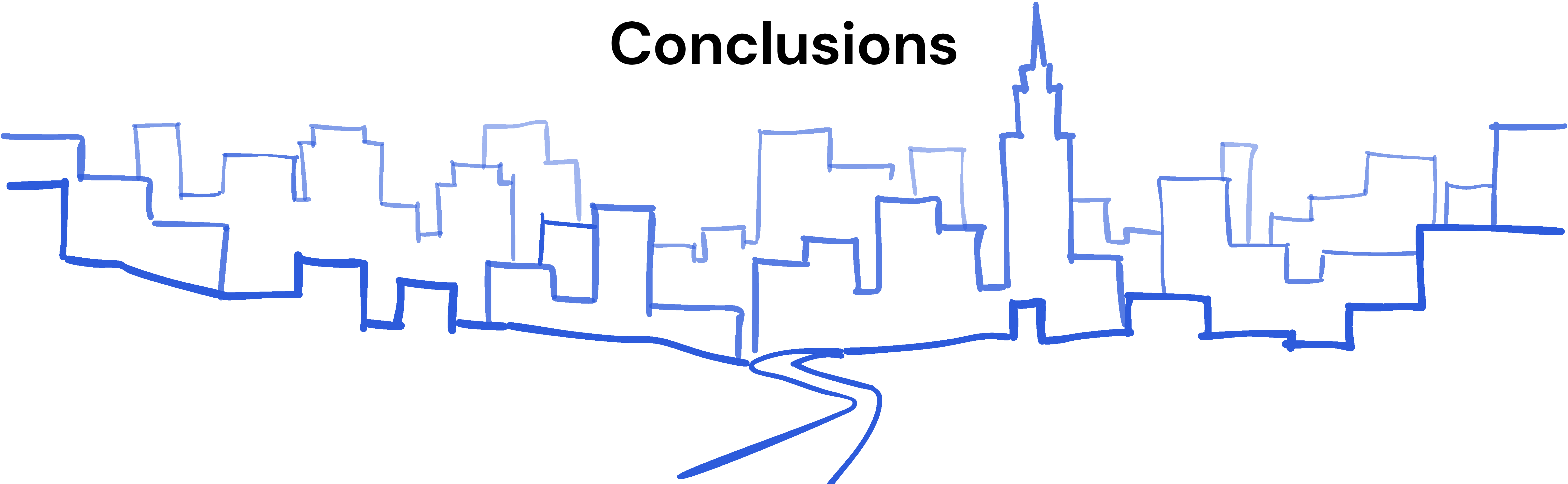


Luckner, M., Łysak, A., & Archanowicz-Kudelska, K. (w druku). Estimating physical activity during travel to school. W Proceedings of the SoGood 2024: ECML-PKDD Workshop on Data Science for Social Good. Wilno, Litwa.



Luckner, M., Łysak, A., & Archanowicz-Kudelska, K. (w druku). Estimating physical activity during travel to school. W Proceedings of the SoGood 2024: ECML-PKDD Workshop on Data Science for Social Good. Wilno, Litwa.

Conclusions



Algorithms: statistical methods, classic machine learning models, deep neural networks

Data generation: AI toolkit usage (download, concatenate, clean data)

Pre-processing: survey data filtering, simulating additional travels, representing data as tables or maps

Enhancement: class balancing, different map coloring

Modeling: travelling time distribution analysis, prediction of walk or car choose, classification of transportation mode

Explanation: feature importance, Shapley values, Grad-CAM, saliency maps