

CZY MOŻEMY PRZEWIDZIEĆ W JAKICH SYTUACJACH MIESZKANKI I MIESZKAŃCY WYBIORĄ POSZCZEGÓLNE ŚRODKI TRANSPORTU?

WYNIKI ANALIZY DLA WARSZAWY

Maciej Grzenda
POLITECHNIKA WARSZAWSKA



Wybór środka transportu

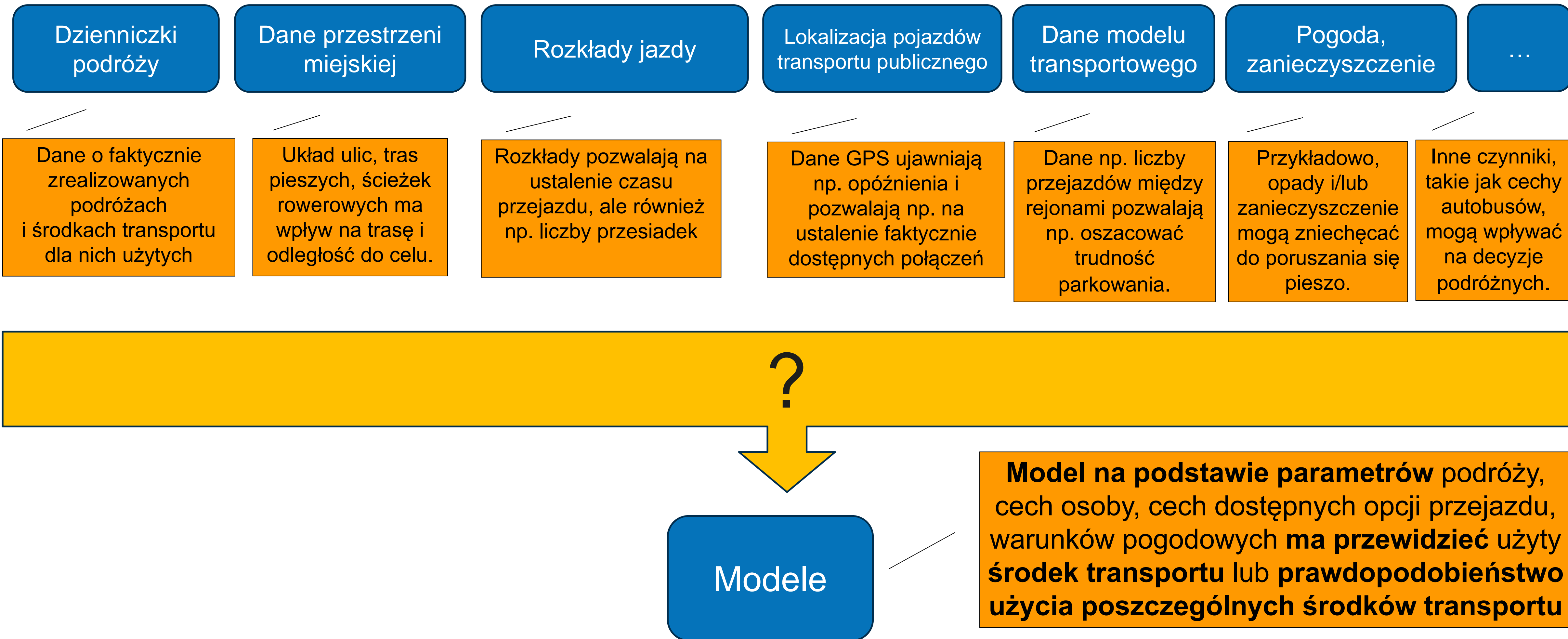
Wstęp

- **Wybór środka transportu** (ang. *travel mode choice*): **wybór metody przemieszczania się w celu realizacji podróży w przestrzeni miejskiej**. Podróż oznacza przemieszczanie się w terenie otwartym – również pieszo i na niewielką odległość.
- **Kluczowe metody przemieszczania się**: użycie transportu publicznego, wykorzystanie samochodu w dyspozycji osoby podróżującej, pieszo, przejazd rowerem
- **Kluczowe pytania**:
 - Czy dla konkretnych cech osoby i podróży możemy skutecznie przewidzieć typowo dobierany środek transportu?
 - **Od czego dokładnie zależą decyzje** mieszkańców?
- Przykłady pytań związanych z postawionym zagadnieniem:
 - Kiedy korzystamy z transportu publicznego? W jakich sytuacjach?
 - Znane czynniki: szybkie połączenia, niewielka odległość do przystanku, warunki pogodowe,...
 - Kiedy uznajemy, że najlepszym rozwiązaniem będzie dotarcie do celu pieszo?
 - Cel naszej podróży jest niedaleko,...
 - Co jednak oznacza „szybkie połączenia”, „cel niedaleko”, „przystanek blisko”? Jak istotne są te i inne czynniki?
- **Zmiennych potencjalnie wpływających na decyzje jest bardzo wiele.**
- Stąd użycie **uczenia maszynowego** (ang. **machine learning**) by stworzyć **modele**, które **ujawnią faktyczne** (a nie deklarowane) **zależności na podstawie analizy danych faktycznych podróży**



Modelowanie wyborów środków transportu

Jak zbudować modele uczenia maszynowego?



Modelowanie wyborów środków transportu

Przygotowanie do modelowania i budowa modeli

Dzienniczki podróży

Dane przestrzeni miejskiej

Rozkłady jazdy

Lokalizacja pojazdów transportu publicznego

Dane modelu transportowego

Pogoda, zanieczyszczenie

...

Platforma informatyczna stworzona w projekcie posłużyła do wyliczenia wartości zmiennych i przygotowania danych gotowych do modelowania.
Dyskusja: w trakcie warsztatów.

Dane o faktycznych podróżach zintegrowane z innymi danymi, w tym dostępnych opcji dotarcia do celu w okresie podróży, w dniach poprzedzających, danymi pogodowymi, cechami urbanistycznymi wokół miejsca podróży i innymi
Osobne zbiory danych dla różnych grup respondentów

Dla każdej grupy mieszkańców powstała grupa modeli predykcyjnych z użyciem kodu stworzonego w projekcie

Platforma Unified Travel Mode Choice Modelling (UTMCM)

PARENTS_W1

PARENTS_W2

CITIZENS_W1_2

Budowa modeli

Budowa modeli

Budowa modeli

Model

Model

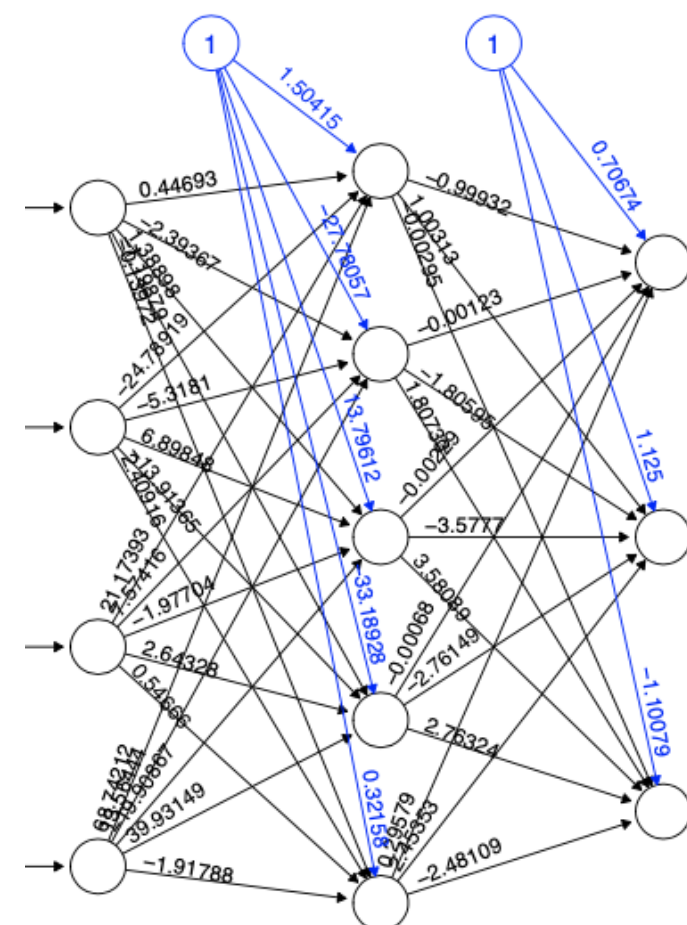
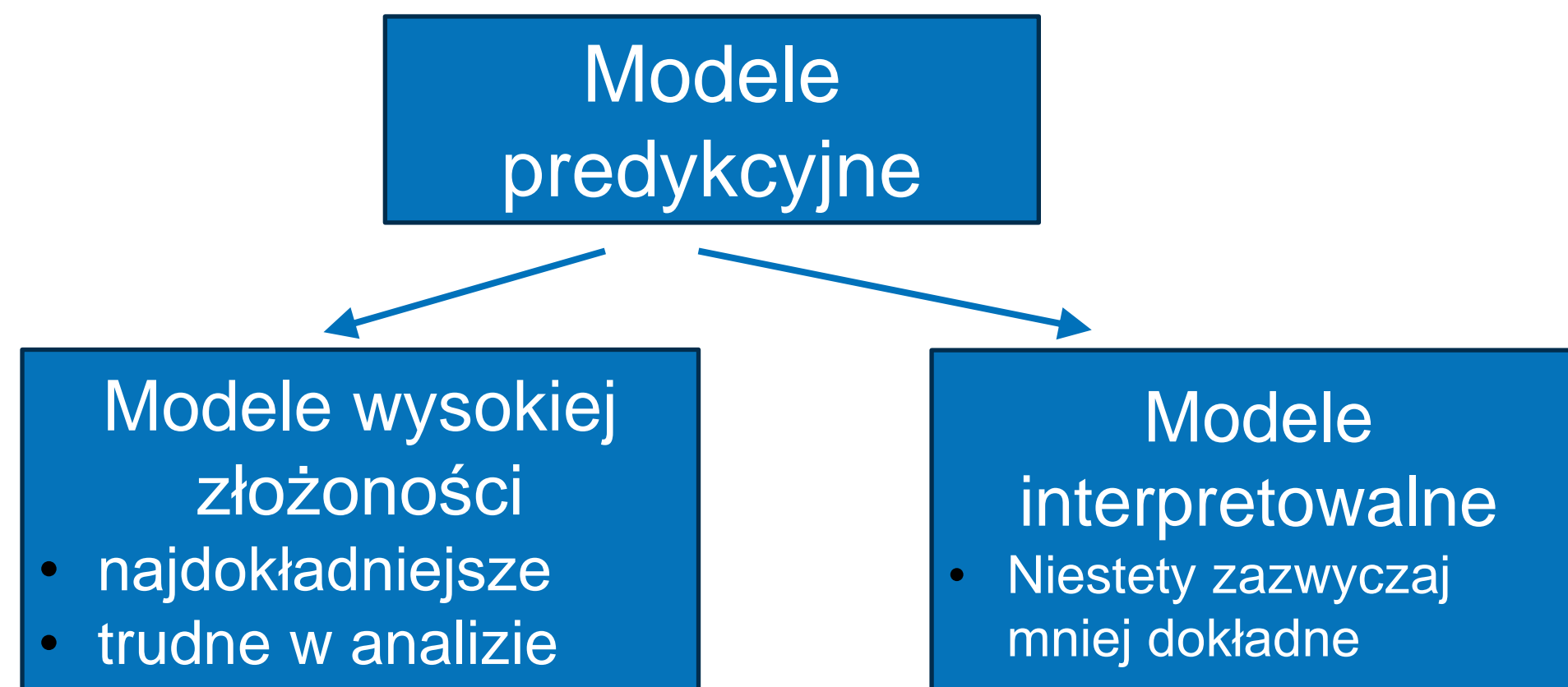
Model

Całość platformy informatycznej i modeli opisywanych w prezentacji została zrealizowana przez zespół PW

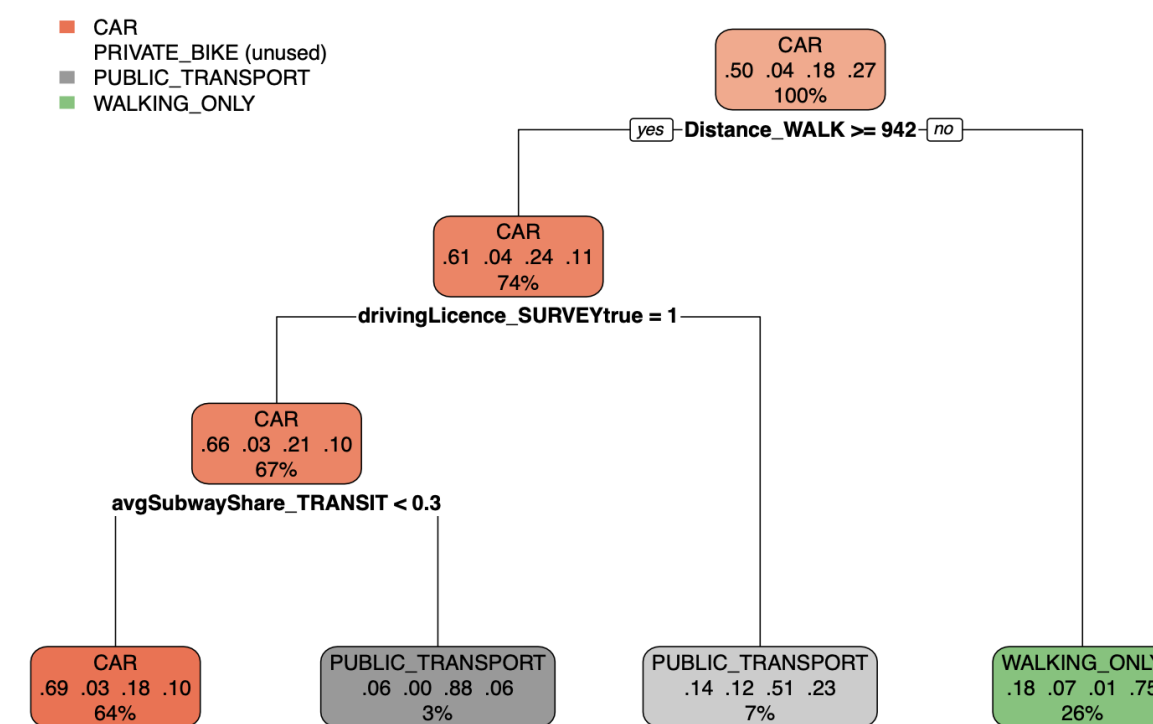


Kategorie modeli uczenia maszynowego

Dlaczego modeli jest wiele?



Rys. 1 Schemat bardzo prostej sieci neuronowej (4 zmienne, predykacja jednej z trzech klas)



Rys. 2 Schemat prostego drzewa decyzyjnego (predykacja jednej z czterech klas)

- Model może bazować na różnych danych, np.:
 - tylko cechy osoby
 - cechy osoby i cechy dostępnych opcji transportu
 - cechy osoby, opcji transportu i warunki pogodowe
 - ...
- W projekcie powstały modele zbudowane dla różnych wariantów - zestawów użytych cech
- W każdym wariancie stworzyliśmy:
 - Modele interpretowalne (drzewa decyzyjne)
 - Modele złożone (np. XGBoost, las losowy, SVM w różnych wariantach, wielowarstwowy perceptron, ...)



Wyniki budowy modeli

Modele złożone

- Modele zostały zbudowane dla rodziców dzieci z trzech szkół, ogółu rodziców i ogółu mieszkańców
- **Modele złożone zgodnie z oczekiwaniami są zazwyczaj najdokładniejsze**
- Użycie tylko danych społeczno-demograficznych nie pozwala na skuteczne przewidywanie wybieranych środków podróży – dokładność predykcji wynosi np. 52.6% dla podróży ogółu mieszkańców Warszawy
- **Połączenie w/w danych i danych wyznaczonych przez platformę, w tym m.in. opcji transportu dostępnych dla danej podróży pozwala poprawnie przewidzieć nawet 70.4% decyzji mieszkańców**
- **Uwzględniając mnogość nieznanymi modelowi czynników** (np. samopoczucie, indywidualne preferencje, stałe lub chwilowe ograniczenia czasowe, posiadanie dedykowanego miejsca parkingowego,) **to wysoka skuteczność**
- Drzewa decyzyjne (modele interpretowalne) mają dokładność zazwyczaj niższą np. dla użycia wszystkich cech 61.3% (-6.6%), 70.4% (b/z), 60.0% (-7.0%)

Zbiór danych	Próba	Dokładność predykcji środka transportu – tylko dane ankietowe	Dokładność predykcji środka transportu – dane ankietowe i wyznaczone przez platformę informatyczną
PAR_W1	Rodzice z wybranych trzech szkół podstawowych	51.3%	67.9%
PAR_W2	Rodzice z różnych szkół podstawowych	58.5%	70.4%
CIT_W1_W2	Ogół mieszkańców Warszawy	52.6%	67.0%

Tabela 1. Dokładność najlepszych modeli przewidywania środka transportu uzyskana na danych testowych (nowsze podróże innych respondentów)



Wyniki budowy modeli

Modele złożone – zmienne kluczowe dla przewidywania środka transportu dla przykładowego modelu



Rys. 3 Najważniejsze zmienne z punktu widzenia wpływu na przewidywanie użycia różnych środków transportu. Model ranger. Dane CITIZENS_W1_2 (ogół mieszkańców Warszawy)

Dla każdej zmiennej poznajemy wpływ na decyzję o użyciu lub nie danego środka transportu. Ta zmienna w największym stopniu wpływa na przejście piesze.

Czas podróży transportem publicznym. Jedna z czterech wartości związanych z progami 20 min, 75 min, 90 min

Odległość od miejsca startu podróży do najbliższej stacji metra. Silny wpływ na użycie ... samochodu i transportu publicznego.

Posiadanie roweru. Istotny wskaźnik użycia ... transportu publicznego.

Obecność dzieci w gospodarstwie domowym. Silny wpływ na użycie samochodu i transportu publicznego. Z innych modeli wynika, że rodzice mają mniejszą tolerancję dla dłuższych spacerów.

Różnica i proporcja czasu przejazdu samochodem (przy założeniu braku zatłoczenia dróg) w stosunku do czasu przejazdu z użyciem transportu publicznego

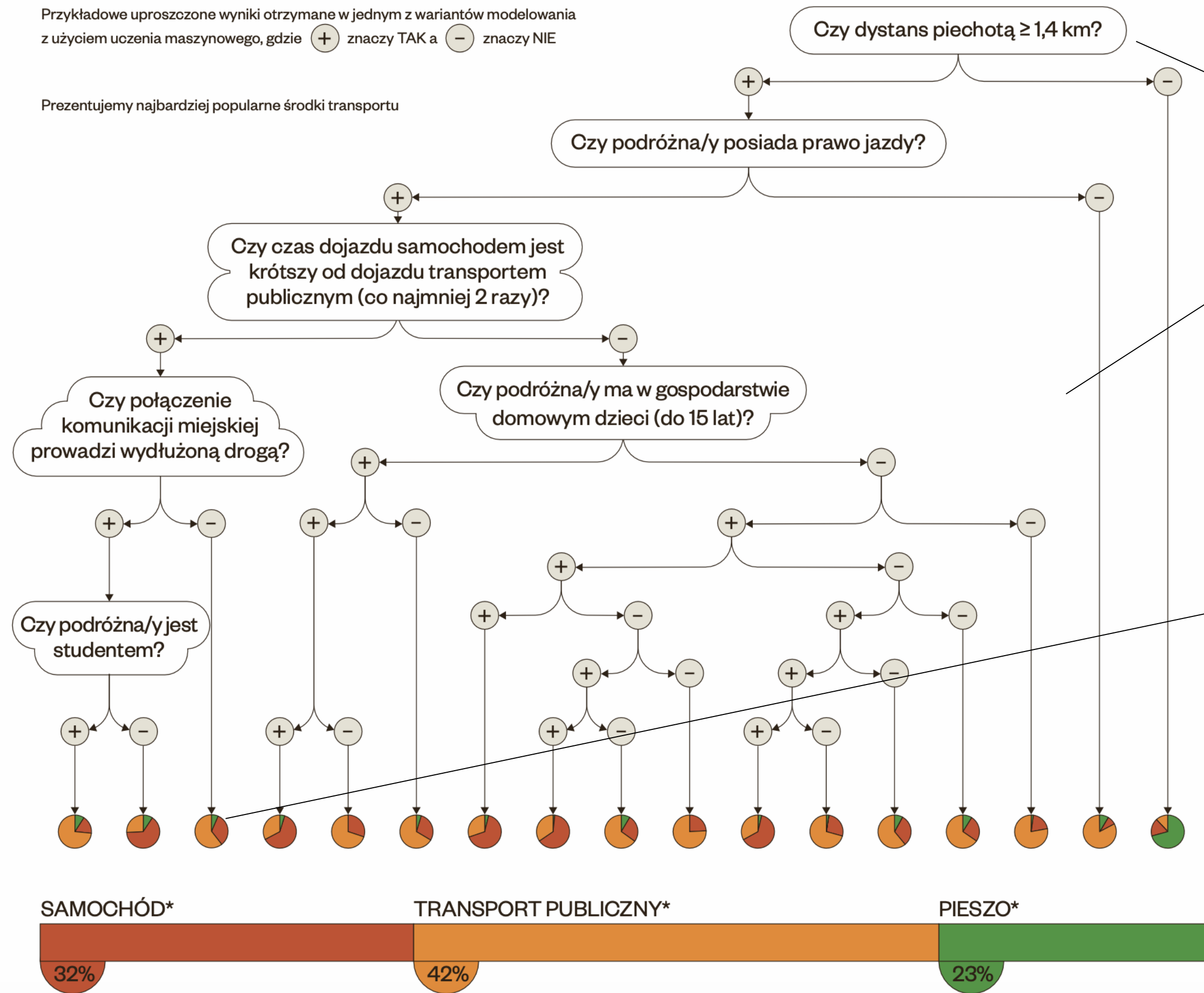


Wyniki budowy modeli

Model interpretowalny – drzewo decyzyjne

Przykładowe uproszczone wyniki otrzymane w jednym z wariantów modelowania z użyciem uczenia maszynowego, gdzie **+** oznacza TAK a **-** oznacza NIE

Prezentujemy najbardziej popularne środki transportu

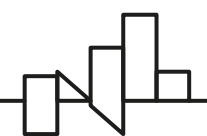


Na infografice przedstawiono uproszczony model typu drzewo decyzyjne. Model przedstawia zależności wykryte w danych faktycznych podróży

W drzewie decyzyjnym w każdym kroku weryfikowane są kolejne warunki, co prowadzi nas do ustalenia najbardziej prawdopodobnego środka podróży i oszacowania prawdopodobieństwa użycia innych.

- Przykład:
- dla podróży na odległość ponad 1.4km
 - realizowanych przez osoby z prawem jazdy
 - gdy czas podróży samochodem jest krótszy co najmniej dwukrotnie od czasu podróży transportem publicznym,
 - ale przejazd prowadzi „prosto do celu”
 - mieszkańcy wybierają zazwyczaj transport publiczny.

Rys. 4 Infografika – uproszczone wybory środka transportu. Mieszkańcy Warszawy

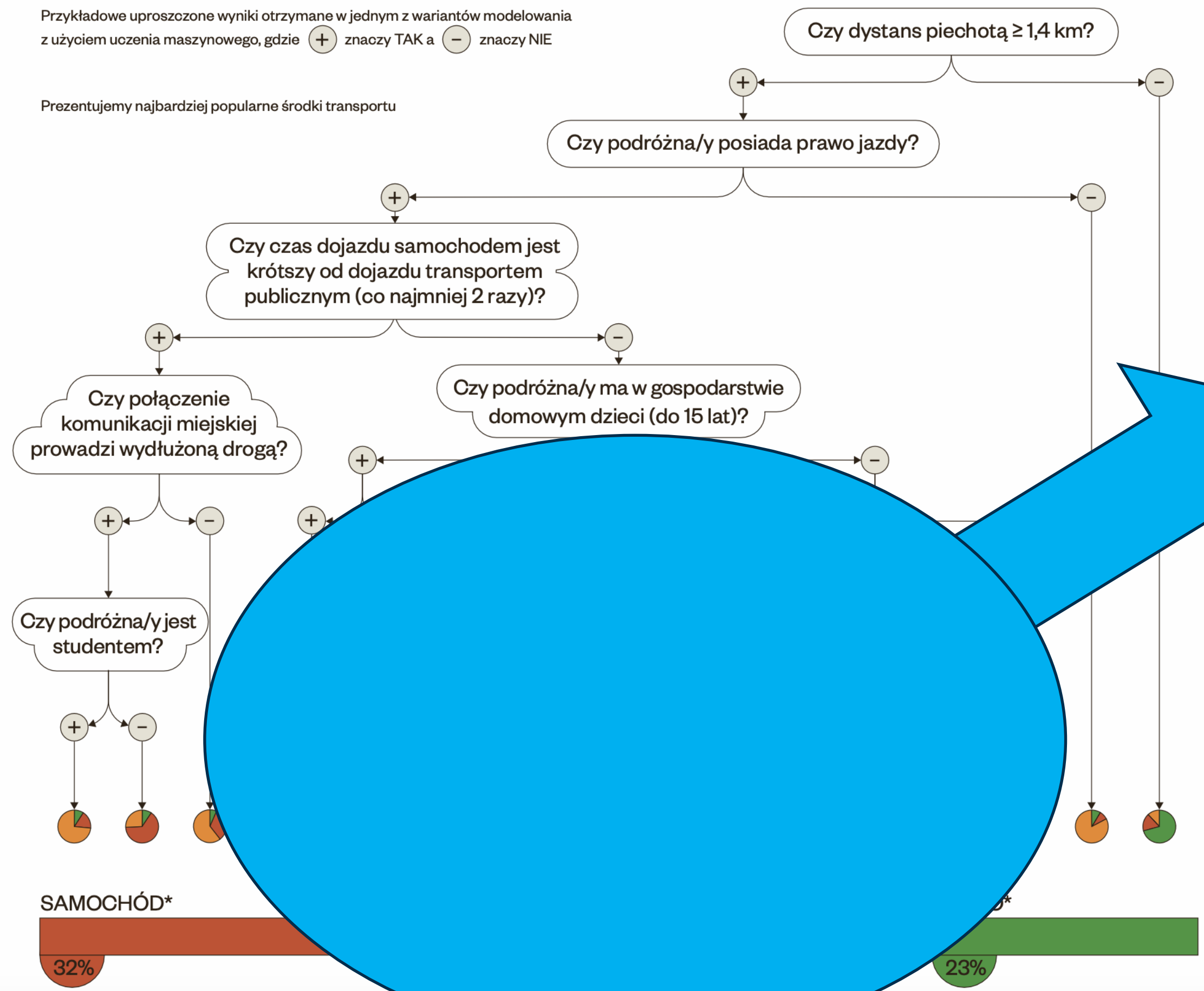


Wyniki budowy modeli

Model interpretowalny – drzewo decyzyjne

Przykładowe uproszczone wyniki otrzymane w jednym z wariantów modelowania z użyciem uczenia maszynowego, gdzie **+** znaczy TAK a **-** znaczy NIE

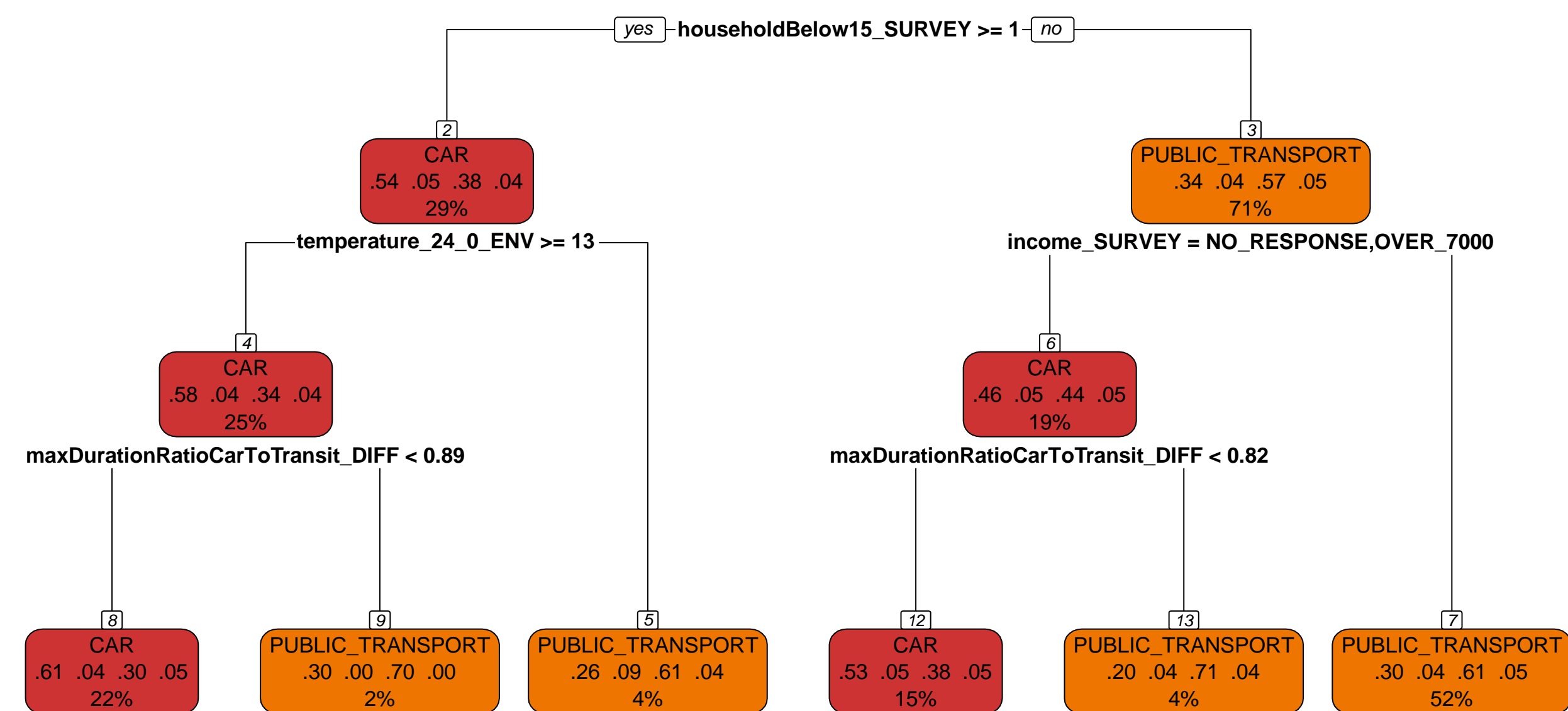
Prezentujemy najbardziej popularne środki transportu



SAMOCHÓD*

32%

23%



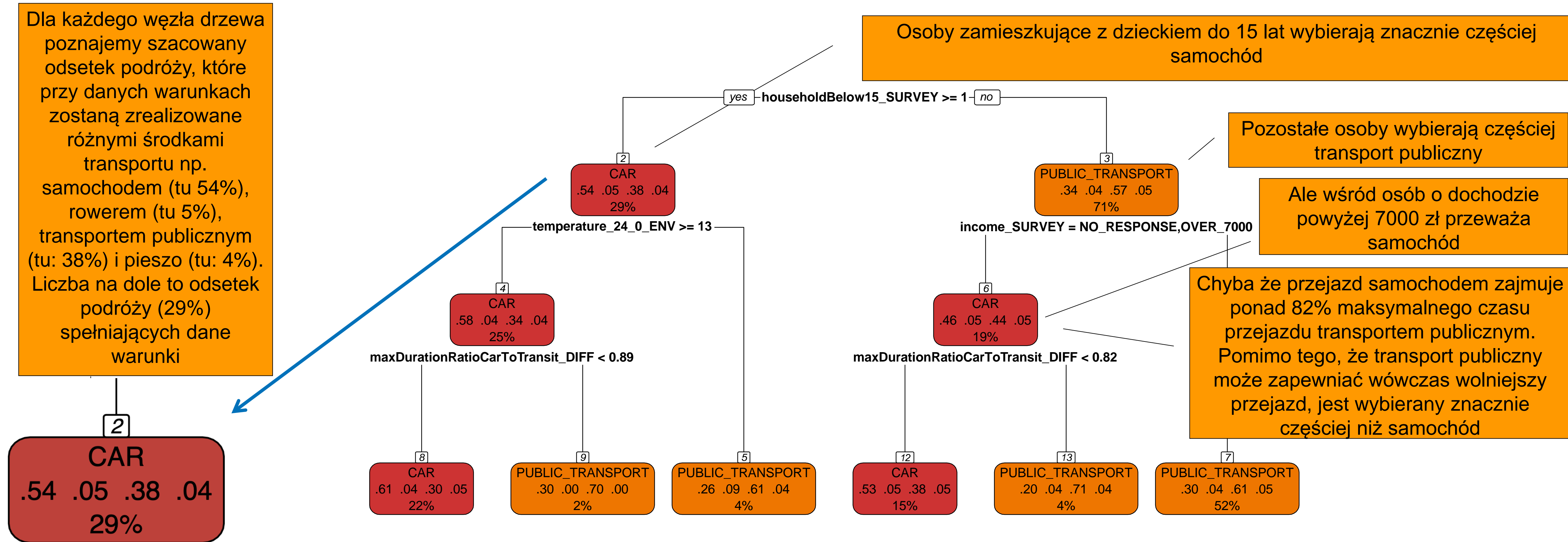
Rys. 5 Drzewo decyzyjne – wybory mieszkańców Warszawy. Przykładowe poddrzewo dotyczy podróży osób posiadających prawo jazdy w przypadku podróży na odległość ponad 1390 m, gdy czas przejazdu samochodem (bez zatłoczenia dróg) jest dłuższy niż 0.47 czasu dotarcia do celu transportem publicznym.

Rys. 5 Infografika – uproszczone wybory środka transportu



Wyniki budowy modeli

Model interpretowalny – pełne drzewo decyzyjne



Rys. 5 Uprozczone drzewo decyzyjne – wybory mieszkańców Warszawy. Drzewo dotyczy podróży osób posiadających prawo jazdy w przypadku podróży na odległość ponad 1390 m, gdy czas przejazdu samochodem jest większy niż około połowa (0.47) czasu dotarcia do celu transportem publicznym



Wybrane pozostałe wyniki

- Przedstawione wyniki stanowią tylko część wyników wykonanych prac
- Inne wyniki:
 - Prezentacja wstępnej wersji systemu i przykładów jego użycia: **Grzenda, M., Luckner, M., Wrona, P. (2023). *Urban Traveller Preference Miner: Modelling Transport Choices with Survey Data Streams*, ECML PKDD 2022, Grenoble**
 - Metoda wykrywania odcinków tras transportu publicznego na których powstają statystycznie istotne opóźnienia i opóźnienia te są redukowane: **Wrona, P., Grzenda, M., Luckner, M. (2022). *Streaming Detection of Significant Delay Changes in Public Transport Systems*, ICCS 2022, Londyn**
 - Metody szacowania trudności parkowania w różnych lokalizacjach i użycia takich metod do poprawy skuteczności przewidywania wyborów środka transportu: **Grzenda, M., Luckner, M., Brzozowski, Ł. (2023). *Quantifying Parking Difficulty with Transport and Prediction Models for Travel Mode Choice Modelling*. ICCS 2023, Praga**
 - Metoda wykrywania zmienności w danych podróży i wyborów mieszkańców oraz adaptacji modeli do zmienności danych: **P. Golik, M. Grzenda, E. Sienkiewicz, *Hybrid Ensemble-Based Travel Mode Prediction***, praca zaakceptowana na konferencję **Symposium on Intelligent Data Analysis, Sztokholm, 24-26 kwiecień 2024**



Podsumowanie

- **Użycie uczenia maszynowego w powiązaniu z wyliczeniem wielu (320 w typowym wariancie) zmiennych pozwala na skuteczne przewidywanie wyborów środka transportu**
- **Szeroka dostępność danych udostępnianych przez m. st. Warszawa** umożliwiła szczególnie ciekawe analizy
- **Efektom prac jest:**
 - **Uzyskanie** zarówno **modeli** złożonych wysokiej dokładności, jak i modeli interpretowalnych
 - Ustalenie rankingu czynników wpływających na podejmowane decyzje
 - **Stworzenie rozwiązań informatycznych umożliwiających dalsze modelowanie we współpracy z ekspertami miast** w tym wspólną analizę i rozwój stworzonych rozwiązań
- Zapraszamy do dyskusji w trakcie i po konferencji oraz współpracy
- Kolejny możliwy krok - przygotowanie kolejnych wariantów modeli uczenia maszynowego bazujących na wybranych zmiennych szczególnie interesujących dla jednostek odpowiedzialnych za transport publiczny

Platforma informatyczna i modele przedstawione w prezentacji zostały opracowane przez zespół Politechniki Warszawskiej w składzie:

- Maciej Grzenda
- Marcin Luckner
- Grzegorz Bagrowski
- Łukasz Brzozowski
- Mateusz Pietrkiewicz
- Adrian Trzeciak
- Przemysław Wrona

Wykorzystano dane ankietowe zgromadzone w projekcie CoMobility pod kierunkiem SGH. Podziękowania dla zespołu projektu CoMobility za współpracę.



Dziękuję za uwagę!

dr hab. inż. Maciej Grzenda, prof. PW
Politechnika Warszawska, Wydział Matematyki i Nauk Informacyjnych (WMI NI)

Maciej.Grzenda@pw.edu.pl

Zapraszam na warsztat „*Rola danych i modeli w zrozumieniu wyborów środków transportu i wpływu tych wyborów na zanieczyszczenia*” w drugim dniu konferencji

przygotowany wspólnie z:

dr hab. inż. Marcin Luckner, prof. PW, Politechnika Warszawska, WMI NI

dr Henrik Grythe, NILU

Marcin.Luckner@pw.edu.pl

heg@nilu.no

www.comobility.edu.pl

