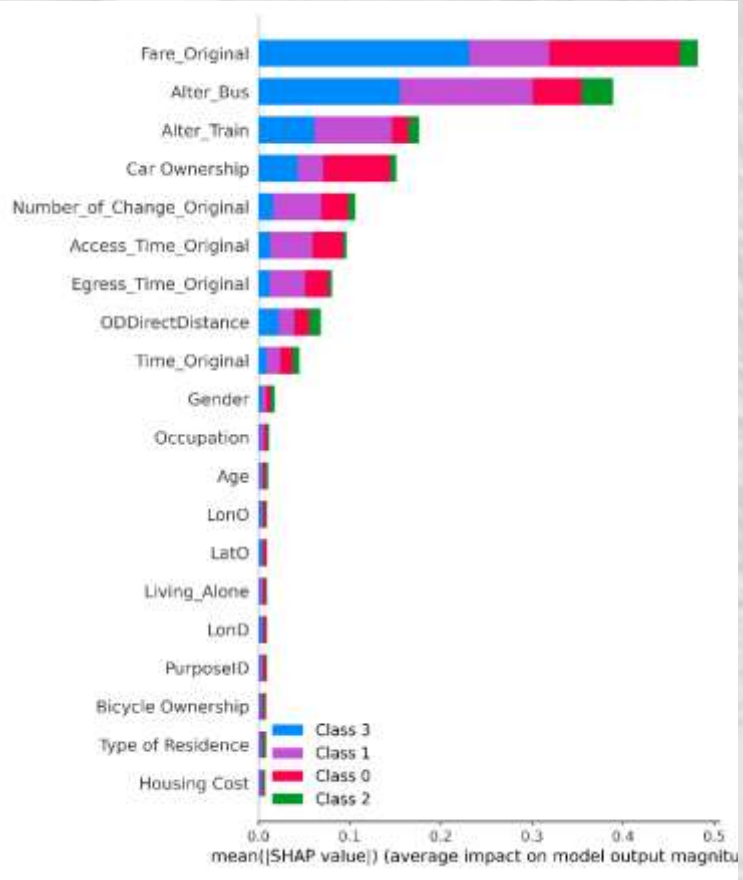
名古屋大学
NAGOYA UNIVERSITY

NU TREND
Nagoya University
TRansport and
ENvironment Dynamics

Accurate Modeling of Travel Mode Choice Behavior Based on Explainable AI



Nagoya University（Yamamoto Miwa Laboratory)

WANG Siyuan. WANG Hongjie. ZHAO Zhijian.

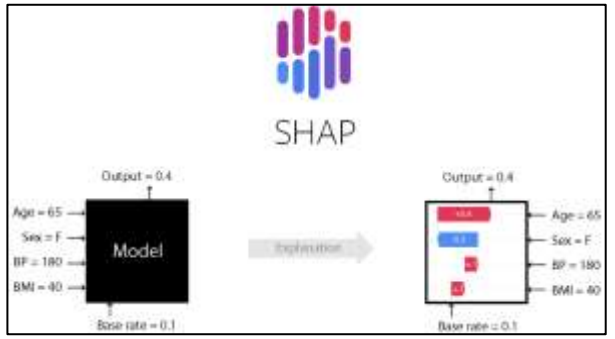
1.Background

1. Challenges and Importance of Mode Choice Behavior Modelling

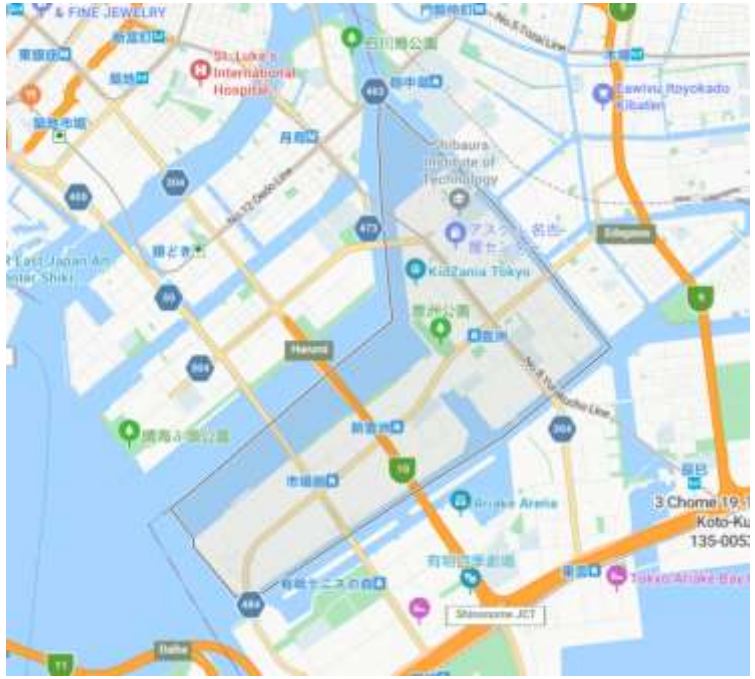
With rapid urbanization, travel mode choice has become a key factor affecting urban traffic efficiency, environmental quality, and sustainable development. Traditional models, such as discrete choice models, struggle to capture complex, nonlinear relationships and the diversity of individual travel behaviors.

2. Emergence of Explainable AI

Recent advances in machine learning and deep learning have improved travel mode choice modeling, but these models are often “black boxes” with limited interpretability. Explainable AI (XAI) addresses this by providing transparency and interpretability, helping researchers and policymakers better understand and trust model predictions.

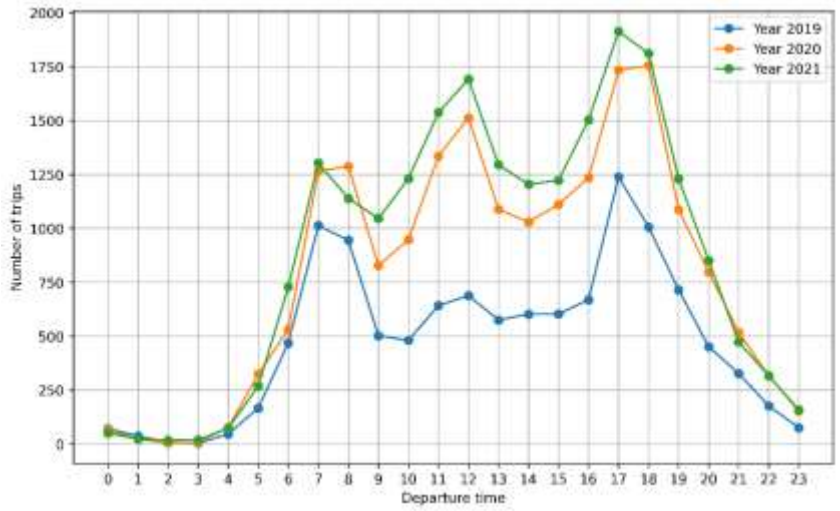


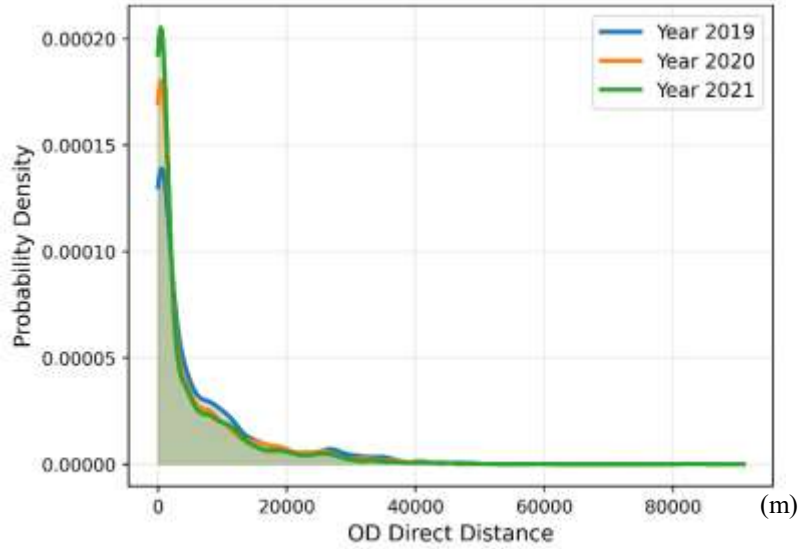
2. Data overview

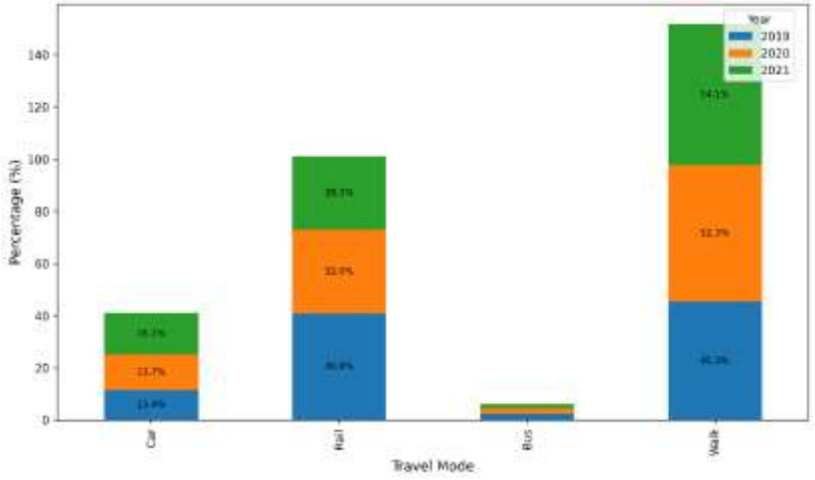


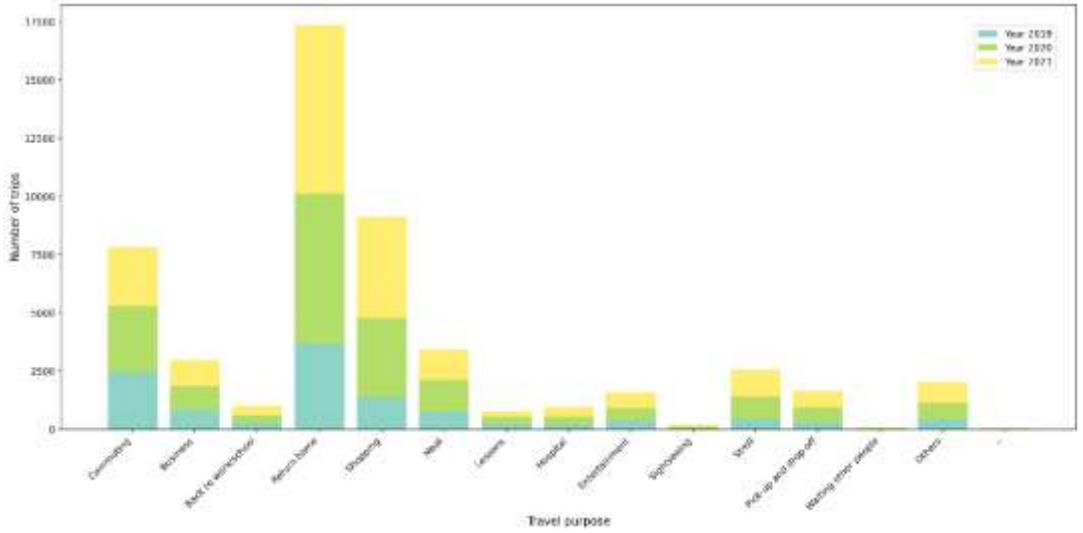
Survey Area	Toyosu
Period	2018~2021
Number of samples	104999 ➔ 51650
Dataset	Toyosu_PP_Survey_Individual_Attribute_Data ; ID_panel_survey ; trip_toyosu

3. Descriptive Analysis

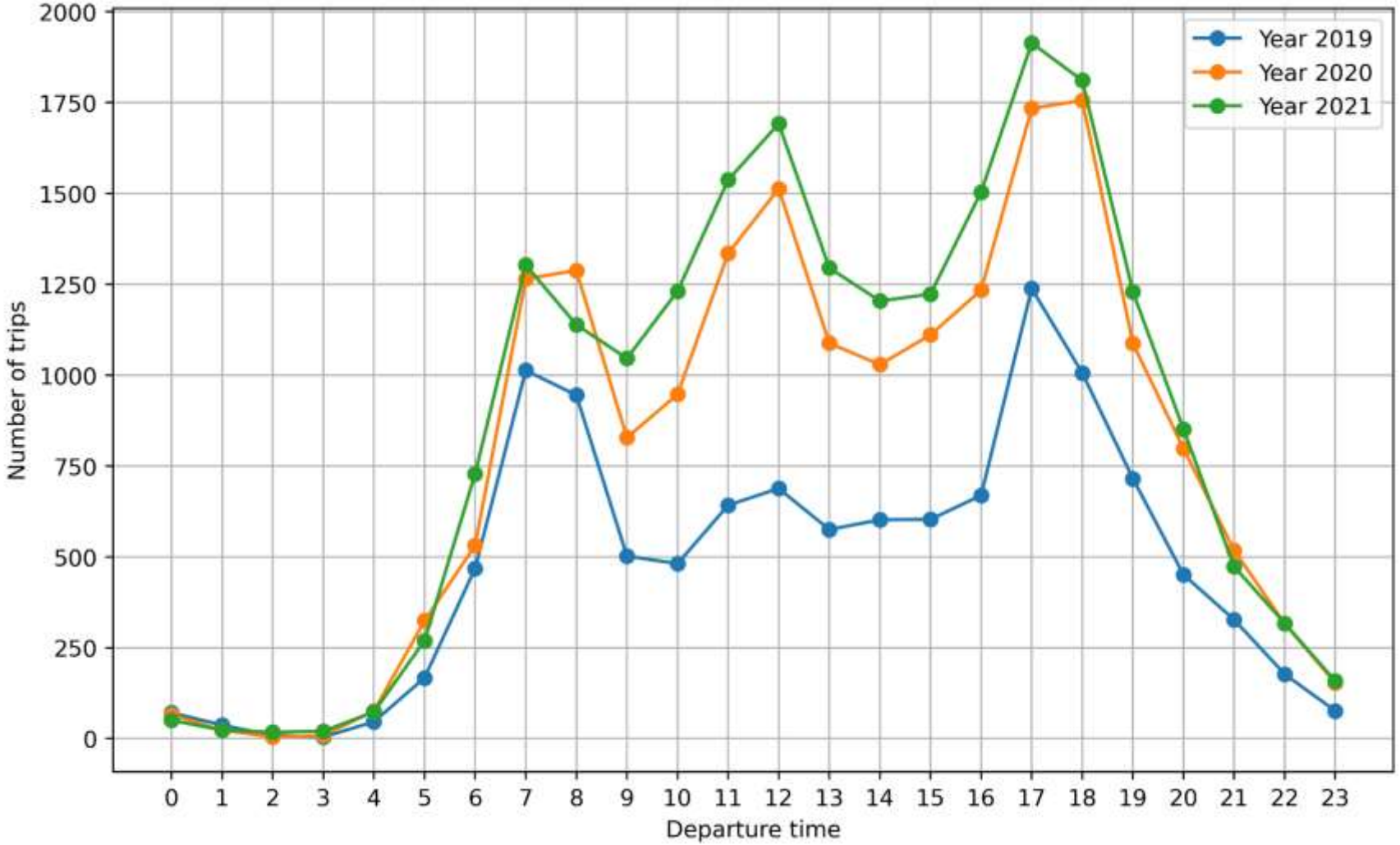




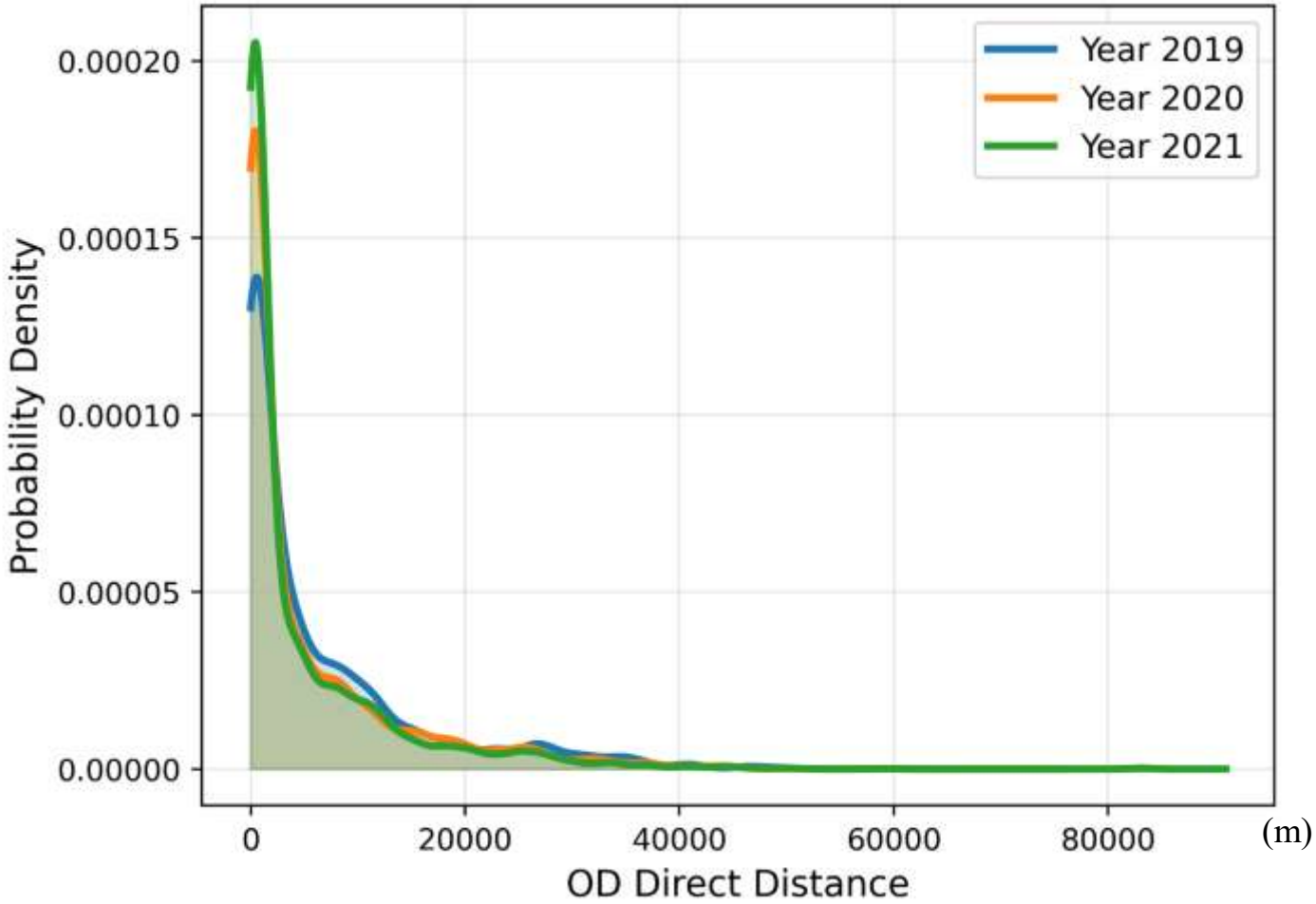




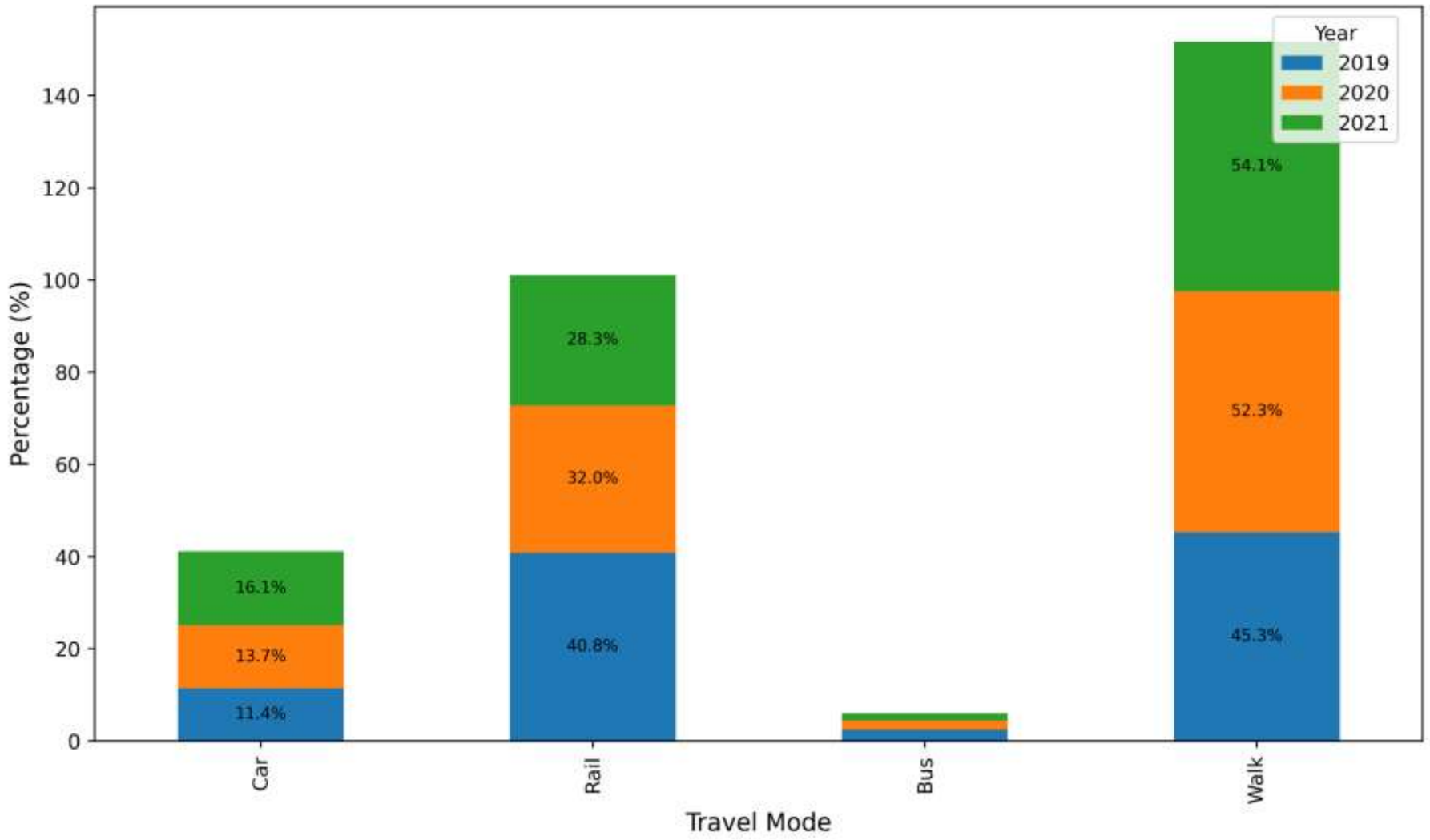
3. Descriptive Analysis



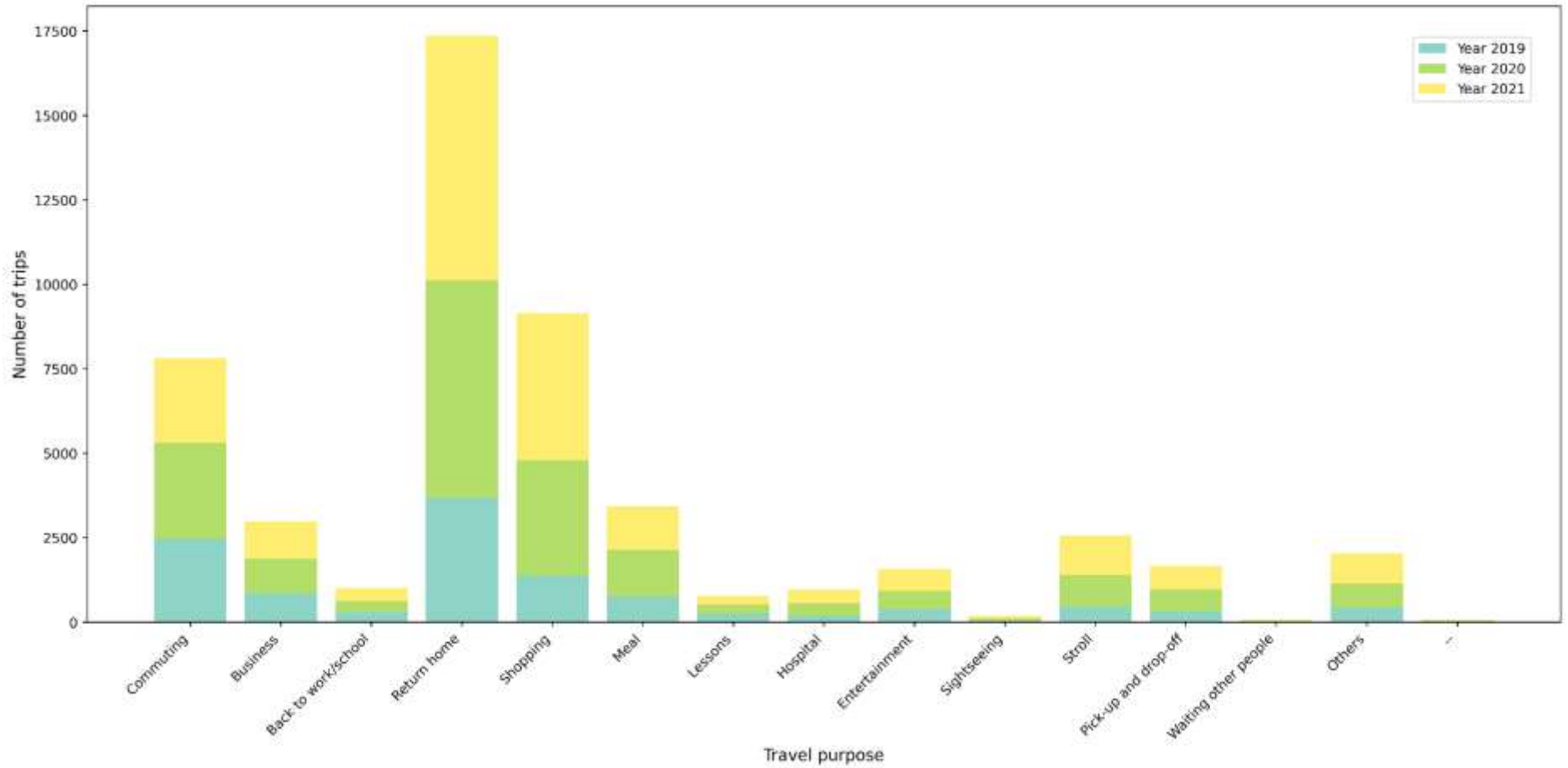
3. Descriptive Analysis



3. Descriptive Analysis



3. Descriptive Analysis

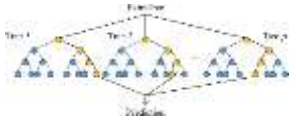


4. Modeling

Dependent Variable:	Mode choice
Independent Variable:	Purpose, Departure time, Year, OD Direct Distance, LonO, LonD, LatO, LatD, Car availiability， Train availiability, Bus availiability, Walk availiability, Gender, Age, Occupation, Income, Type of Residence, Years of Residence, ……

4. Modeling

Machine learning models



ExtraTree

Random Forest (RF)

Lipid remodeling regulator 1 (LRL1)

FLAML

Microsoft's AutoML toolkit

Grid search

Model training

hyperparameter tuning

Goal:

Max f1_macro score

Best fitness

Macro-F1 score=0.8864

v.s

Traditional method

Logistic regression model

Macro-F1 score= 0.7703

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

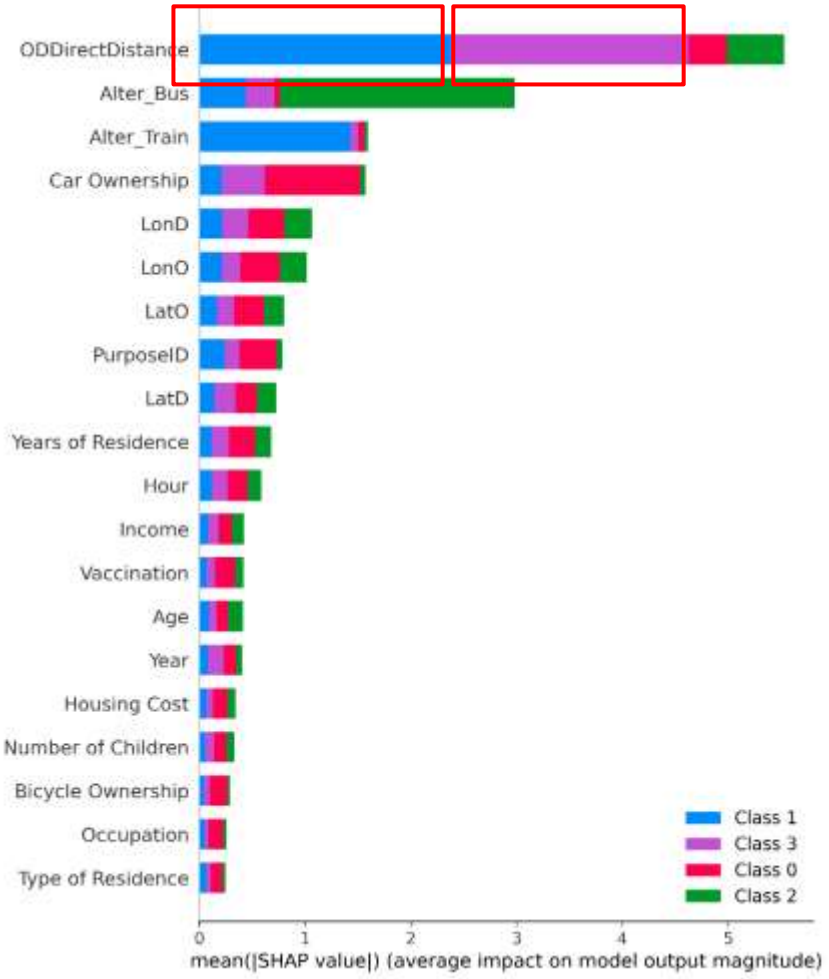
$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + RecalL}$$

$$F_1 - \text{macro} = \frac{F_{1(\text{class0})} + F_{1(\text{class1})} + F_{1(\text{class2})} + F_{1(\text{class3})}}{4}$$

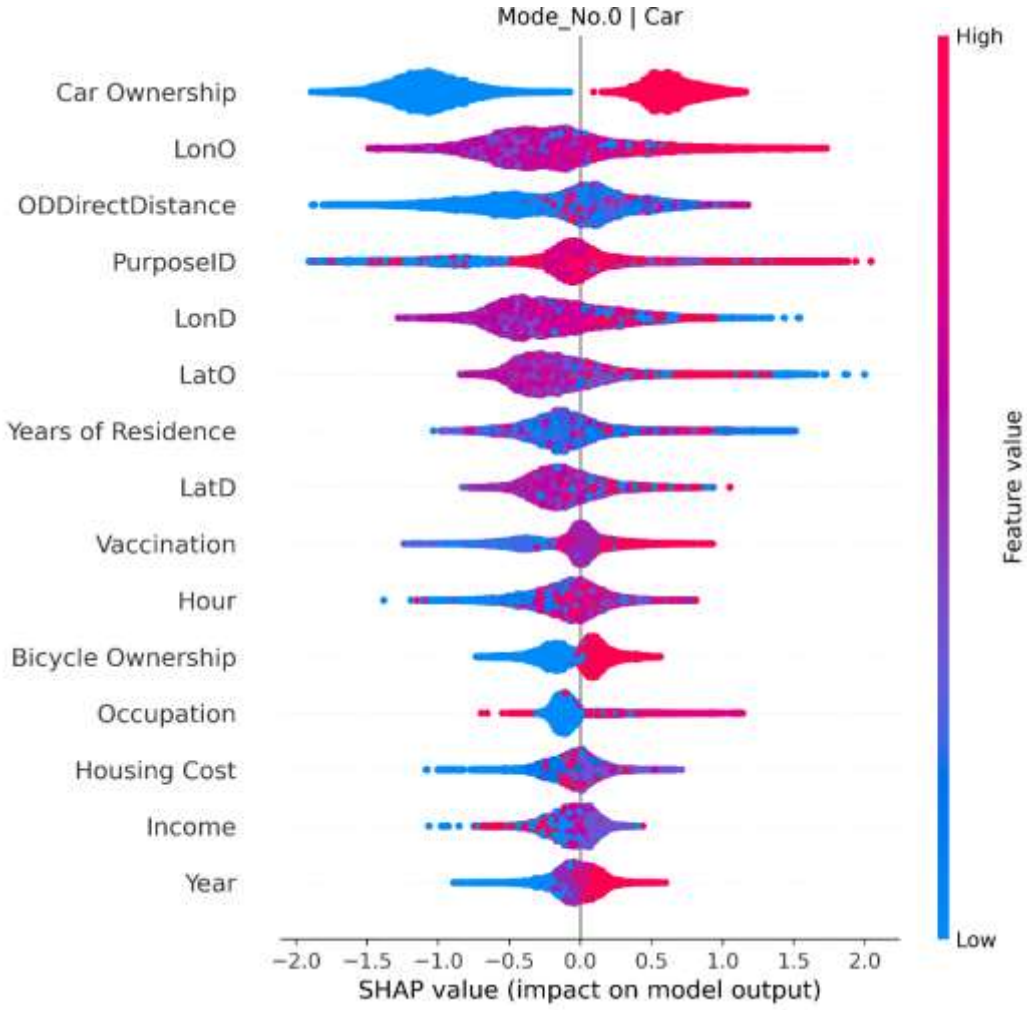
5. Result

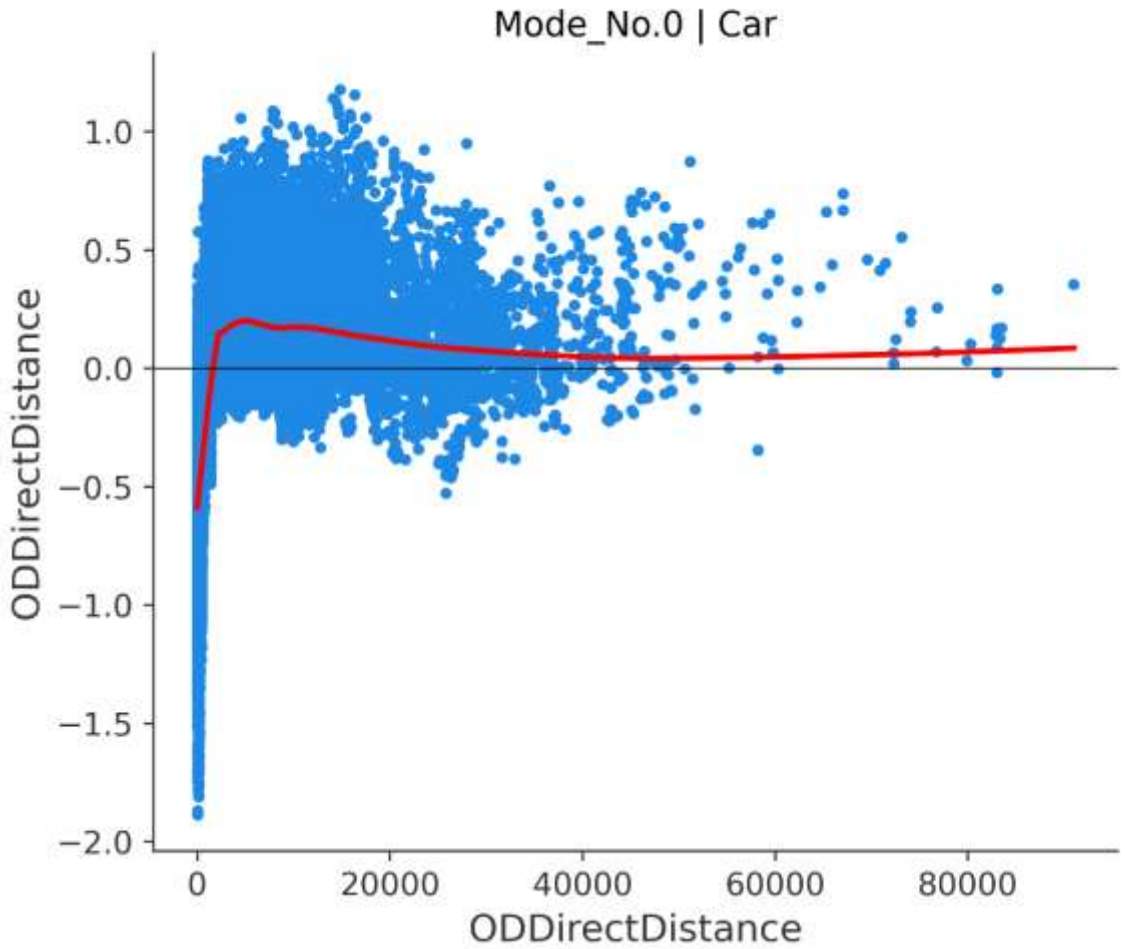


Class 0 --- car Class 1 --- rail
Class 2 --- bus Class 3 --- walk

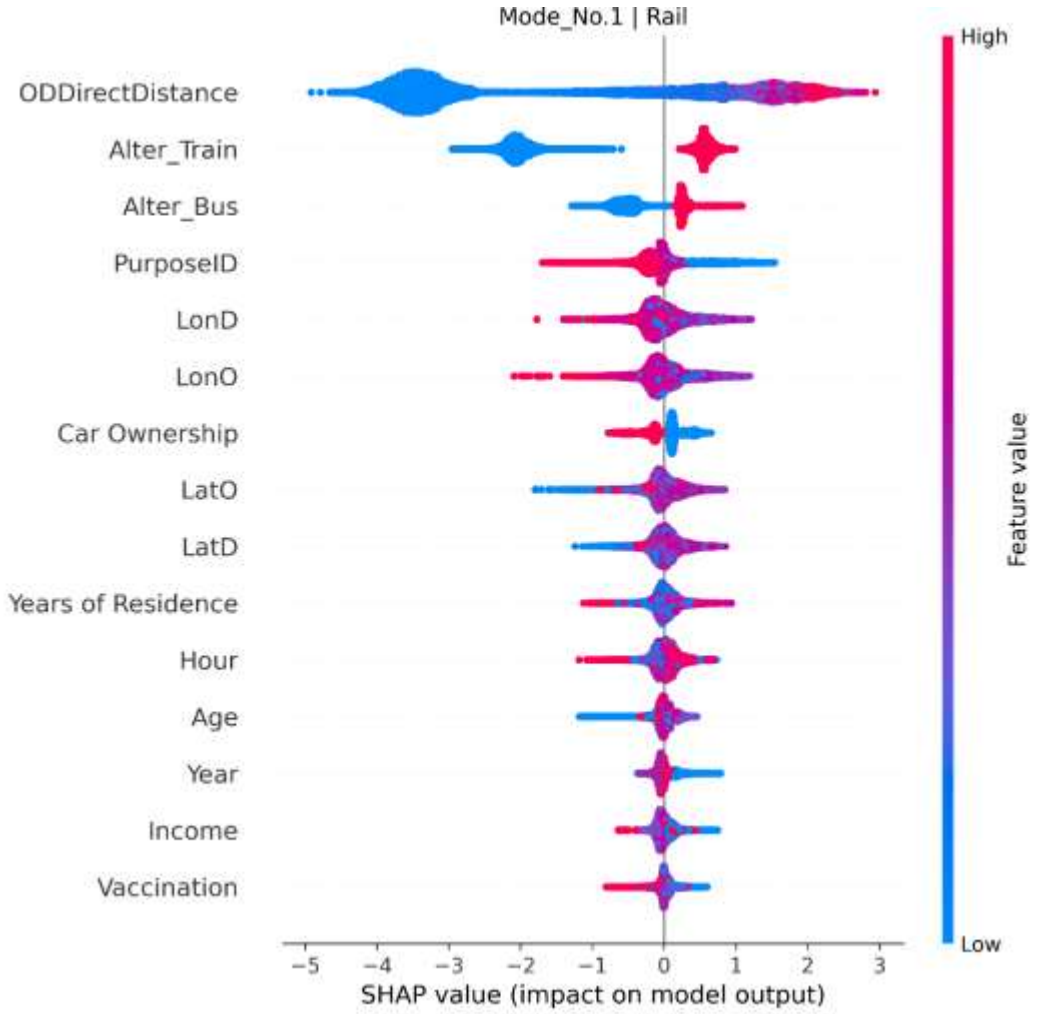
- For the rail or walk, the “OD direct distance” has the biggest importance.
- For the bus, the “Alter_bus”, which means the availiability to use bus, has the biggest importance.
- Car ownership has the greatest impact on whether to choose a car.

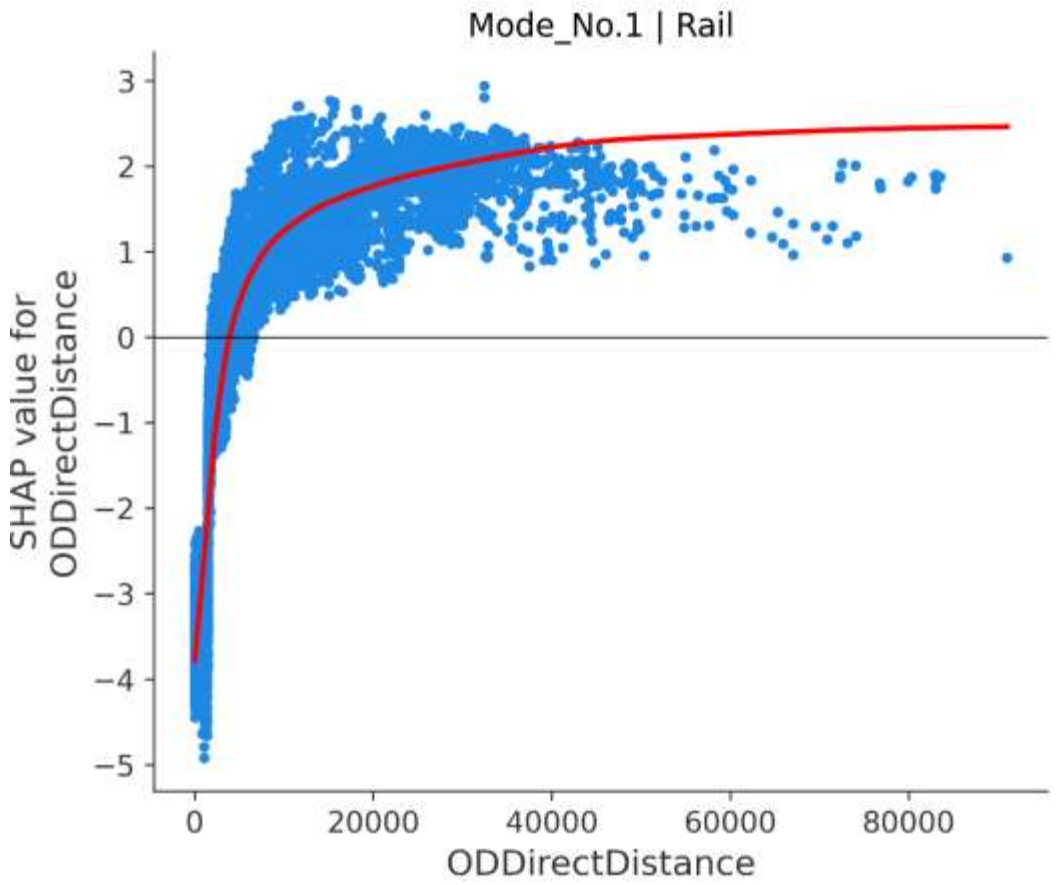
5. Result


Mode_No.0 | Car
Car Ownership
LonO
ODDirectDistance
PurposeID
LonD
LatO
Years of Residence
LatD
Vaccination
Hour
Bicycle Ownership
Occupation
Housing Cost
Income
Year
−2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0 1.5 2.0
SHAP value (impact on model output)
High
Feature value
Low

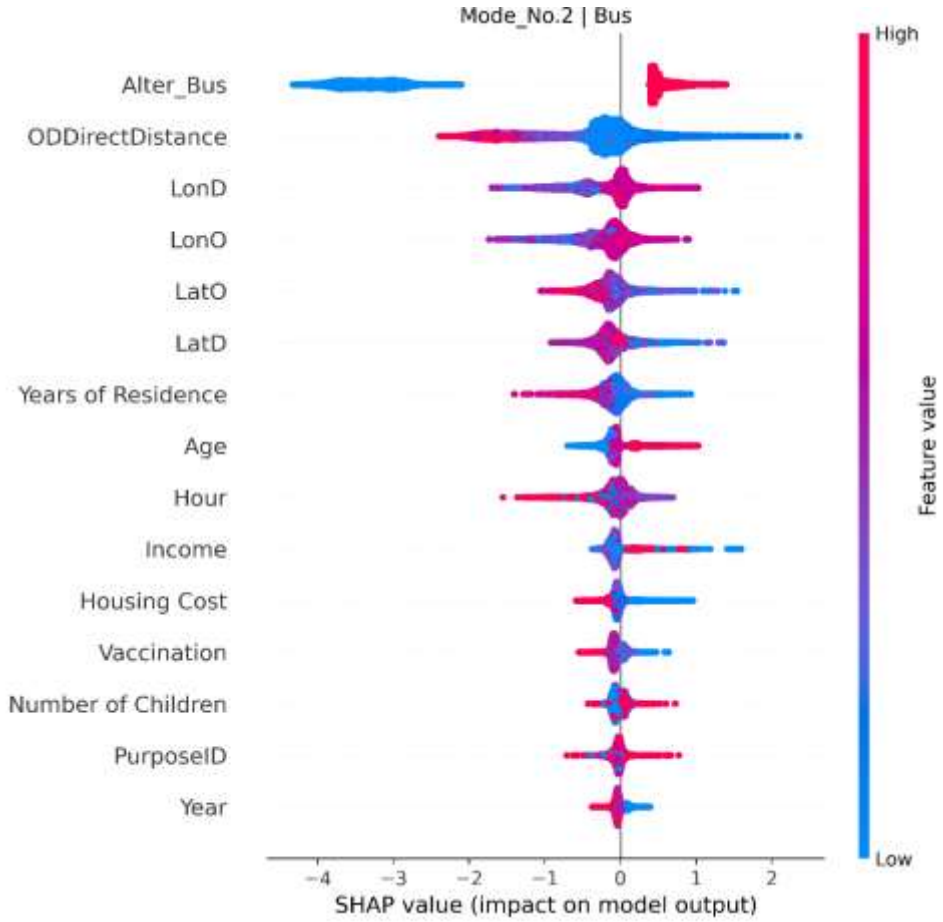

Mode_No.0 | Car
ODDirectDistance
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
20000
40000
60000
80000
ODDirectDistance

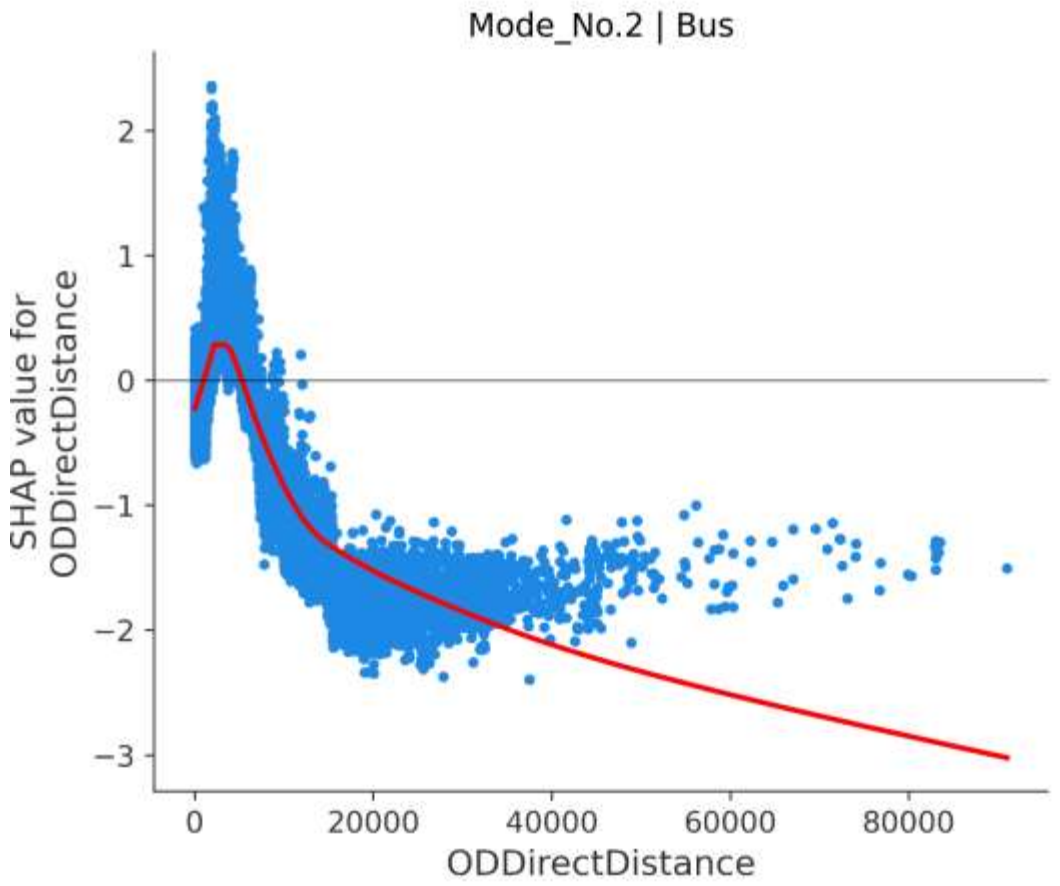
5. Result


Mode_No.1 | Rail
ODDirectDistance
Alter_Train
Alter_Bus
PurposeID
LonD
LonO
Car Ownership
LatO
LatD
Years of Residence
Hour
Age
Year
Income
Vaccination
−5 −4 −3 −2 −1 0 1 2 3
SHAP value (impact on model output)
High
Feature value
Low

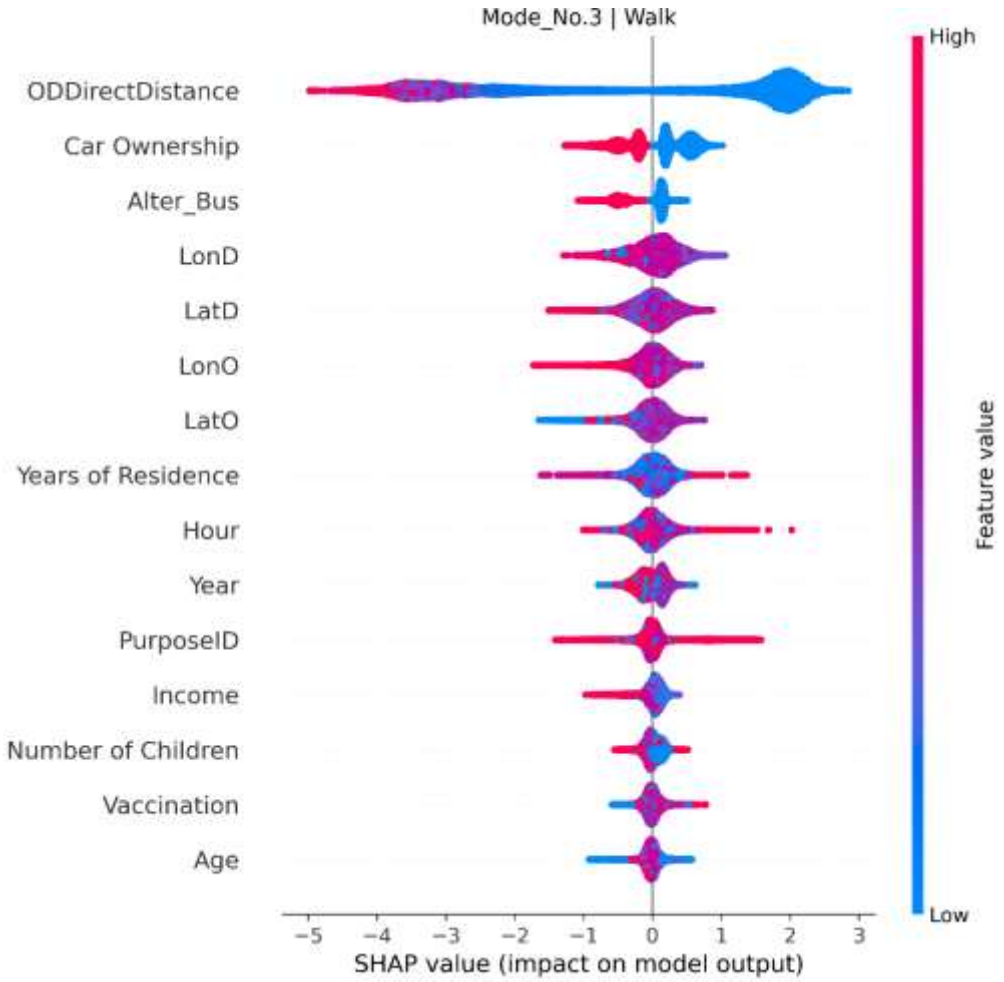

Mode_No.1 | Rail
SHAP value for ODDirectDistance
3
2
1
0
−1
−2
−3
−4
−5
0
20000
40000
60000
80000
ODDirectDistance

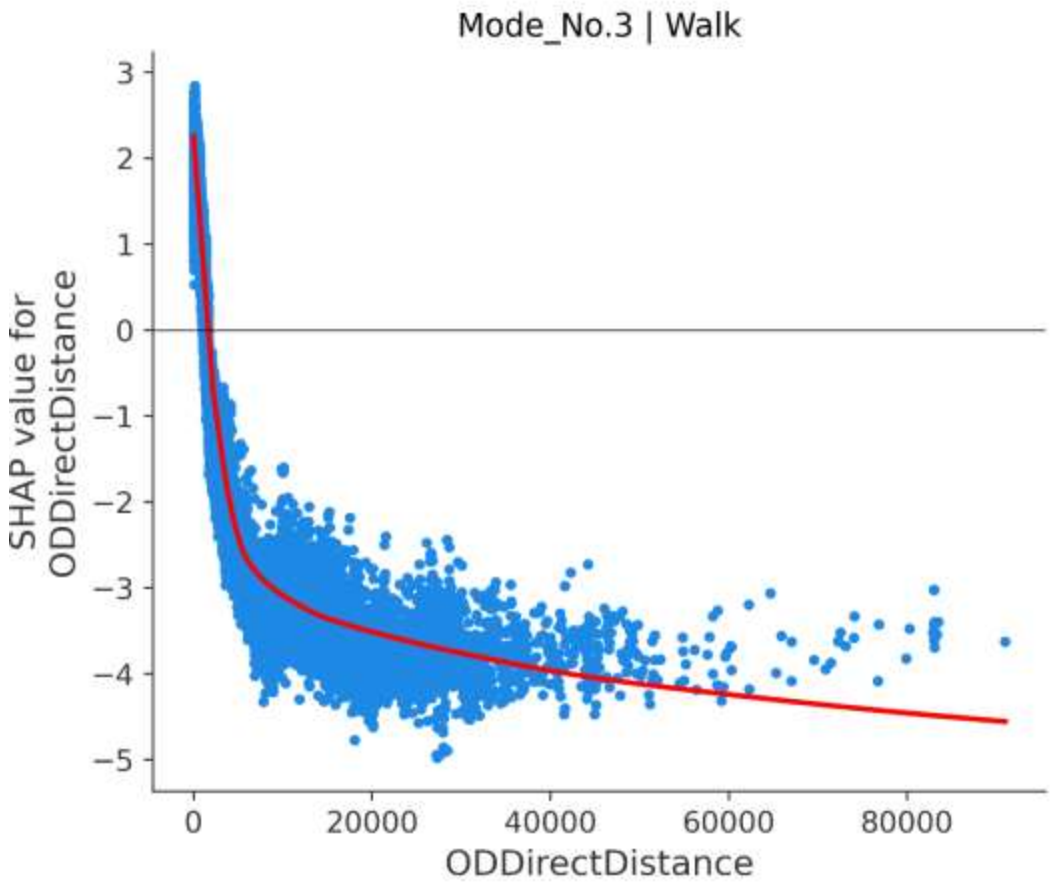
5. Result


Mode_No.2 | Bus
Alter_Bus
ODDirectDistance
LonD
LonO
LatO
LatD
Years of Residence
Age
Hour
Income
Housing Cost
Vaccination
Number of Children
PurposeID
Year
−4
−3
−2
−1
0
1
2
SHAP value (impact on model output)
High
Feature value
Low


Mode_No.2 | Bus
SHAP value for ODDirectDistance
2
1
0
−1
−2
−3
0
20000
40000
60000
80000
ODDirectDistance

5. Result


Mode_No.3 | Walk
ODDirectDistance
Car Ownership
Alter_Bus
LonD
LatD
LonO
LatO
Years of Residence
Hour
Year
PurposeID
Income
Number of Children
Vaccination
Age
−5 −4 −3 −2 −1 0 1 2 3
SHAP value (impact on model output)
High
Feature value
Low


Mode_No.3 | Walk
SHAP value for ODDirectDistance
3 2 1 0 −1 −2 −3 −4 −5
0 20000 40000 60000 80000
ODDirectDistance

6. Summary

- This study applied several Explainable AI (XAI) methods to improve the modeling of travel mode choice behavior.
- Compared to traditional logistic regression, machine learning models achieved higher accuracy and macro-F1 scores. (XGBoost's Macro-F1 score=0.8864)
- Key influencing factors were identified:

Rail & walk: OD direct distance

Bus: Bus availability (Alter_bus)

Car: Car ownership

- Use the SHAP to analyze the factors, allowing better understanding of model decisions.

名古屋大学
NAGOYA UNIVERSITY

NU TREND
Nagoya University
TRansport and
ENvironment Dynamics

Thanks for your listening.

WANG Siyuan. WANG Hongjie. ZHAO Zhijian.