Shared micro-mobility choice modeling using random forest

1 Introduction

Shared micro-mobility services have been popular worldwide in the past few years, including dockless escooters, dockless and docked bikes and e-bikes. It is important to understand their usage and related influencing factors for better urban planning and transportation management. In this project, you will use random forest to solve a shared micro-mobility choice modeling problem.

2 Aim

The overall aim is to predict users' choice of shared micro-mobility services at the trip level based on the related influencing factors (e.g., points of interest, weather, time) using random forest classification. Zurich is chosen as the study area, in Figure 1. Three types of shared micro-mobility services are considered in this project, including docked bike, docked e-bike, and dockless e-scooter.

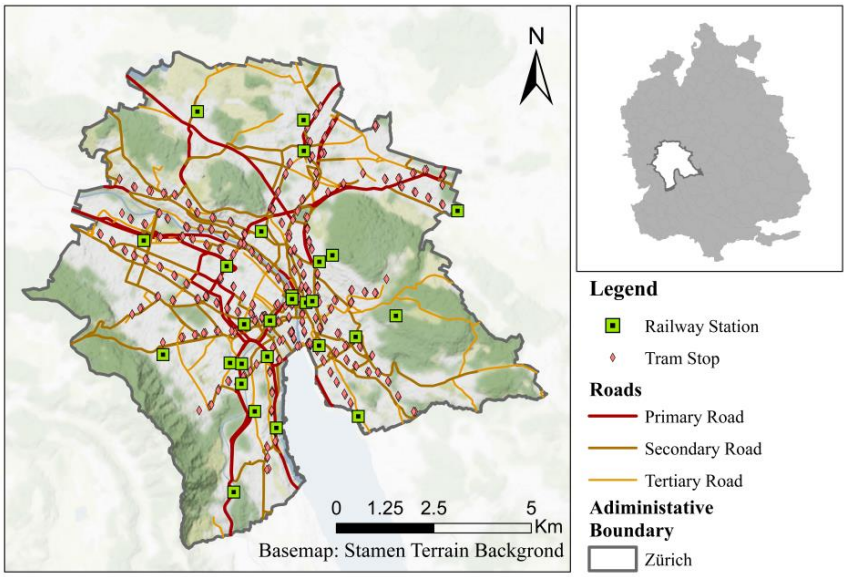

Figure 1. Study area (Ren et al., 2023).

3 Data

The following data has been prepared and can be used for the implementation of the project. The data called 'micromobility_choice.csv' is available. The csv file includes 15,000 records and 20 columns, which cover the target and features:

'id' column stores the id for each trip.

Target/label: 'TripType' column storing the type of shared micro-mobility service used for each trip. 1: bike, 2:e-bike, 4: e-scooter.

Features: duration, distance, elevation_difference, temperature, windSpeed, day_week, hour_day, restaurant, public facility, medical, accommodation, tourist attraction, transportation, education. Table 1 shows the description on the features.

'slng', 'slat', 'elng', and 'elat' represent the longitude and latitude of the start and end points for each trip.

Table 1. The description on the features.

Features	**Description**
duration	Trip duration
distance	Trip distance
elevation_difference	The difference in elevation at the start and end
temperature	Weather-related factors
windSpeed	Weather-related factors
day_week	Whether the trip happened on a weekday or weekend
hour_day	The start time to each hour of the day
restaurant	The number of related POI surrounding the trip end
public facility	The number of related POI surrounding the trip end
medical	The number of related POI surrounding the trip end
accommodation	The number of related POI surrounding the trip end
tourist attraction	The number of related POI surrounding the trip end
transportation	The number of related POI surrounding the trip end
education	The number of related POI surrounding the trip end

4. Workflow

Based on the provided data, the following tasks will be completed. The ratio of training and test data can be 70%:30% in task 1 and 2. Calculate the evaluation metrics: accuracy, precision, recall and f1-score to evaluate the prediction performance.

Task 1: Develop random forest classification model using scikit-learn python library to predict the choice based on the features. In this task, you can keep the hyperparameters on default settings.

Task 2: Implement hyperparameter tuning and develop the optimal random forest classification model. Compare the prediction performance with that of the two models in Task 1.

Task 3: Conduct k-fold cross validation and calculate the average evaluation metrics: accuracy, precision, recall and f1-score.

Task 4: Based on the model in Task 2, visualize the start and end points for the misclassified trips in maps. The visualization for the three types of shared micro-mobility services should be implemented separately.

Write a report within 5 pages to show and describe the results.

5. Oral presentation

Your project must be presented orally. Each group should give a presentation for about 15 minutes to explain the aim, data, tasks and the results.