Univerzitet u Sarajevu
Elektrotehnički fakultet
Sarajevo

UNiVERSiTé MARiE & LOUiS PASTEUR
UFR STGI

Eco-friendly Carpool Optimization Using Clustering and Travel Data

Students: Amina Hromić, Nerma Kadrić. Nadina Miralem and Azra Žunić

Our project aims to develop an optimization model for carpooling that reduces CO_2 emissions and travel costs by grouping passengers based on travel similarity. Using clustering algorithms like KMeans and AGNES with Gower's distance on multi-dimensional travel data, we identify efficient carpool groups. We further optimize transportation modes by substituting car trips with faster and greener train rides where possible.

Data Sources: Simulated datasets including travel missions, passenger profiles, and vehicle CO_2 emission rates.

Features: Included travel start/end locations, trip types, vehicle capacity, travel time, distance, cost, return trip indicator and emission factors.

The global architecture of our system is designed as a streamlined data processing pipeline, illustrating the key phases from raw input data to the final optimized carpooling results (see Figure 1).

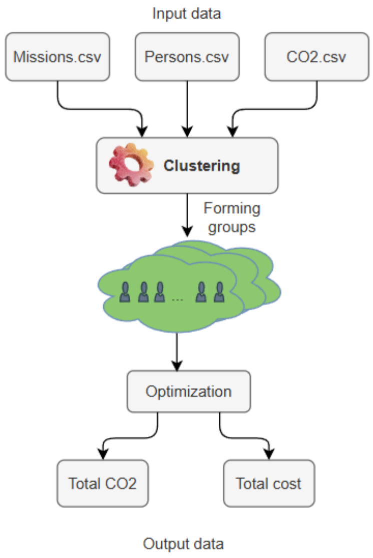

Figure 1: Overview of the system architecture for clustering-based carpool optimization and emissions reduction.

At the core, the system integrates three primary input datasets: missions.csv (trip details), persons.csv (passenger profiles), and co2.csv (vehicle-specific CO_2 emission rates). These heterogeneous datasets are merged and undergo a thorough preprocessing stage involving feature encoding, scaling, and filtering to prepare the data for subsequent analysis.

The methodology involved preprocessing the data by applying one-hot encoding to categorical variables and scaling the numerical features to prepare for clustering. To reduce dimensionality and better visualize the clustering structure, Principal Component Analysis (PCA) was also applied. Two clustering algorithms were tested: KMeans and AGNES. The effectiveness of the clusters was assessed using silhouette scores, which provided insight into the cohesion and separation of the identified groups. Furthermore, all clustering algorithms were applied with a different number of features from the dataset. Experiment one was done with only four features, including: travel distance, cost of the trip, CO_2 emission per kilometer, is_return_trip (categorical variable indicating round trips). As for experiment two, eleven features in total were used and the results compared afterwards. These eleven features included: start_city, end_city, travel_type, vehicle_type, has_car, car_capacity, start_hour, km, total_cost, co2_per_km and is_return_trip.

Following clustering, an optimization step was performed to form carpool groups that respected vehicle capacity constraints and considered temporal proximity, specifically within a ±2-day window. Additionally, a mode substitution strategy was applied, where car trips were replaced with train rides whenever train travel offered a faster alternative, further reducing emissions.

The results showed that the best balance between cost reduction and CO_2 emission minimization was achieved with the AGNES algorithm using 30 clusters in experiment two with 11 features. This configuration resulted in approximately a 61% reduction in both cost and emissions compared to a baseline greedy approach. Despite these promising findings, the average carpool group size remained relatively small, around two people per group, indicating there is potential for further optimization to increase group sizes and improve efficiency. Looking forward, the project faces several challenges and opportunities for enhancement. Increasing the average size of carpool groups without compromising convenience or cost is a key focus. Incorporating real-time traffic data and user preferences could make the model more dynamic and adaptable. Additionally, extending the system to support real-time grouping and regrouping would provide practical benefits in everyday mobility scenarios.

Deliverables:

- Code repository with data preprocessing, clustering, and optimization scripts.
- Visualizations of cluster formations and evaluation metrics inside GitHub repository
- Final report and one-slide presentation.
- Developer's guide
- Final presentation of the project