Transport Demand Prediction Using Regression Models

Submitted By: Manoj Choudhary

Supervised By: Meghana Gowda

Date: September 2025

Problem Statement

- Predict the number of seats Mobiticket can expect to sell per ride.
- Accurate demand prediction helps optimize fleet allocation and pricing.
- Improves efficiency and reduces empty seats.

Objectives

- Build regression models using historical booking data.
- Evaluate Random Forest and XGBoost models.
- Provide insights into factors affecting seat demand.

Dataset Overview

- Total Records: 51,645 bookings
- Features: 10 (ride_id, seat_number, payment_method, travel_date, travel_time, travel_from, travel_to, car_type, max_capacity)
- Target Variable: seats_sold (aggregated per ride)

Exploratory Data Analysis

- Most rides sold between 20–45 seats.
- Higher bookings observed on weekdays (Mon–Fri).
- Morning rides showed greater demand than late-night rides.
- Route and day of week strongly influence demand.

Data Preprocessing

- Aggregated seat bookings per ride to compute seats_sold.
- Extracted temporal features: day of week, month, hour.
- Encoded categorical variables (route, payment_method, car_type).
- Dataset was clean with no missing values.

Models Used

- Random Forest Regressor: Ensemble of decision trees, robust and interpretable.
- XGBoost Regressor: Gradient boosting algorithm, efficient and powerful.

Results & Evaluation

- Evaluation Metrics: RMSE, MAE, R^2.
- Random Forest: RMSE ≈ 5.120, MAE ≈ 3.450, R^2 ≈ 0.890.
- XGBoost: RMSE ≈ 5.050, MAE ≈ 3.400, R^2 ≈ 0.900.
- Both models performed well, Random Forest showed strong interpretability.

Key Insights

- Day of Week and Route are the strongest predictors of seat demand.
- Max Capacity also strongly influences number of seats sold.
- Random Forest and XGBoost produced reliable forecasts.

Conclusion & Future Scope

- Historical data can effectively predict transport demand.
- Helps improve planning, pricing, and resource allocation.
- Future Work: include holidays/events, weather data, time-series models, hyperparameter tuning.