

Executive Summary

Objective: predict transport delay and explain drivers.

- RandomForest achieved higher R^2 ; LinearRegression used for interpretability.
- Top predictors: scheduled time, route frequency, weather severity.
- Winsorization applied at 1%/99% to reduce outlier impact.

Recommendations:

1. Validate on larger dataset
2. Use temporal CV
3. Explore causal analysis