

Traffic Congestion Forecasting — Model Evaluation & Refinement

Author: J Navya Bhargavi

Component: Data Modelling & Model Evaluation

Objective:

Develop predictive models to forecast hourly traffic volumes at road junctions based on historical data. The aim is to enhance traffic management and reduce congestion.

1. Model Development & Training

Models Implemented:

- Gradient Boosting Regressor (baseline robust model)
- XGBoost Regressor (if available)
- LSTM (for sequential dependency modeling, optional)

Feature Engineering:

- Temporal features: hour, day of week, month, weekend flag
- Lag features: 1 hour, 24 hours, 1 week
- Rolling statistics: rolling mean and std for 3h, 24h, 168h windows
- Weather and event features (temperature, precipitation, event flag)

Data Split:

- Time-based 80/20 split (training: earliest 80%, validation: most recent 20%)
- Cross-validation using TimeSeriesSplit (5 folds)

2. Evaluation Metrics

Metrics used:

- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)
- R-squared (R^2)

Example (synthetic dataset results):

Junction	Model	MAE	RMSE	R^2
101	Gradient Boosting	17.4	22.6	0.94
102	Gradient Boosting	18.2	23.1	0.93
103	Gradient Boosting	20.1	25.7	0.91

Interpretation:

- Lower MAE/RMSE indicate strong predictive accuracy.
- High R^2 (>0.9) confirms the models capture major temporal patterns.

3. Model Validation & Cross-Validation

Validation Approach:

- Time-based split ensures the model is only evaluated on unseen future data.
- k-fold TimeSeriesSplit ($k=5$) evaluates model generalization.