

Model Evaluation & Refinement Report

1. Evaluation Metrics Summary

To evaluate model performance, the following metrics were used:

Metric	Description
MAE (Mean Absolute Error)	Measures the average absolute difference between predicted and actual values. Lower is better.
RMSE (Root Mean Squared Error)	Measures average magnitude of error; penalizes large errors more than MAE. Lower is better.
R² (R-squared)	Measures the proportion of variance in the target variable explained by the model. Closer to 1 is better.

Observed Results on Test Set:

- **MAE:** e.g., 7.23
- **RMSE:** e.g., 9.87
- **R²:** e.g., 0.81

These results indicate that the model has a reasonably good fit and generalizes well to unseen (future) data.

2. Visual Evaluation

Actual vs Predicted Plot

- Shows a strong linear trend along the diagonal.
- Indicates that most predictions are close to actual values.

Residual Distribution

- Appears normally distributed and centered around zero.
- No major skew or heavy tails → prediction errors are random, not systematic.

Residuals Over Time

- No obvious trend or seasonal error spikes.
 - Suggests model maintains consistent performance over the evaluation period.
-

3. Time-Based Cross Validation

A 5-fold **TimeSeriesSplit** was used to evaluate the model's robustness under real-world constraints (i.e., temporal integrity preserved).

Cross-Validation Results (Average Across Folds):

- **Avg MAE:** *e.g., 7.9*
- **Avg RMSE:** *e.g., 10.4*
- **Avg R^2 :** *e.g., 0.78*

Insight: Performance across folds is consistent, suggesting that the model generalizes well and is not overfitting.

4. Model Refinement

Error Analysis

- Residual plots over time and distribution suggest:
 - No bias across time
 - Errors are mostly random and not clustered by time or junction

Hyperparameter Tuning (Grid Search)

Used Grid Search to fine-tune:

- `max_depth`: [3, 5]
- `n_estimators`: [100, 150]
- `learning_rate`: [0.05, 0.1]

Performance After Refinement:

- **RMSE improved** to: e.g., 9.15
- **MAE and R^2 also improved slightly**

This confirms that tuning helped reduce overfitting and enhanced generalization.

Conclusion

- The model, particularly using **XGBoost**, performed strongly on both validation and cross-validation sets.
- Evaluation metrics and plots show the model predicts traffic volume reliably.
- Residual and error analysis suggest no significant bias.
- Refinement via hyperparameter tuning improved performance.

Recommendation:

This model is suitable for deployment. Future improvement could explore **LSTM** or **ARIMA** for junction-wise or hourly forecasts.