Report on Model Evaluation & Refinement

1. Introduction

After developing predictive models for hourly traffic congestion, it is essential to evaluate their performance and refine them to improve accuracy and reliability. This phase focuses on measuring prediction quality, identifying model weaknesses, and iteratively improving model performance. The evaluation and refinement process ensures that the final model generalizes well and performs reliably under real-world conditions.

2. Evaluation Metrics

To assess model performance, the following regression metrics were used:

- **Mean Absolute Error (MAE):**
  Measures the average absolute difference between predicted and actual traffic values. It provides an intuitive understanding of prediction error.

- **Root Mean Squared Error (RMSE):**
  Penalizes larger errors more heavily, making it suitable for identifying poor performance during peak traffic hours.

- **R^2 (Coefficient of Determination):**
  Indicates the proportion of variance in traffic volume explained by the model.

These metrics align with the project goal of minimizing traffic prediction error while maintaining model interpretability.

3. Model Performance Evaluation

Two primary models were evaluated:

- **ARIMA (Baseline Time-Series Model)**

- **Gradient Boosting Regressor (Tree-Based Model)**

Evaluation was performed using a time-based validation split, ensuring that training data consisted of past observations and validation data represented future traffic conditions.

Key Observations

- The ARIMA model captured general temporal trends but struggled to model the influence of external factors such as weather and events.

- The Gradient Boosting model achieved lower MAE and RMSE values and higher R^2, indicating better predictive performance.
- Visual inspection using actual vs predicted plots showed that Gradient Boosting closely followed real traffic patterns compared to ARIMA.

4. Diagnostic Analysis

To understand model errors, detailed diagnostic analyses were performed:

4.1 Bias Analysis

Residuals were analyzed during identified peak hours. The results showed that the initial Gradient Boosting model tended to underestimate traffic during peak congestion periods, indicating mild bias.

4.2 Variance Analysis

Residual scatter plots revealed moderate variance, particularly during off-peak hours, suggesting sensitivity to noise in certain time periods.

4.3 Error Pattern Analysis

Hourly residual analysis indicated systematic errors during peak hours, while off-peak periods exhibited more random noise. These findings motivated further refinement of temporal features.

5. Model Refinement

Based on diagnostic insights, iterative improvements were applied:

5.1 Feature Refinement

Additional temporal features were introduced:

- **24-hour lag feature** to capture daily repetition in traffic patterns
- **6-hour rolling average** to smooth short-term fluctuations

Redundant and less informative features were reduced to prevent overfitting.

5.2 Algorithm Enhancement

The model was upgraded from standard Gradient Boosting to a Histogram-based Gradient Boosting Regressor, which:

- Handles missing values more robustly
- Improves computational efficiency
- Enhances generalization on tabular data

5.3 Re-Training and Re-Evaluation

The refined model was retrained using the enhanced feature set and re-evaluated using the same validation strategy. The refined model achieved:

- Reduced MAE and RMSE
- Improved R^2 score
- More stable performance across time-based cross-validation folds

6. Cross-Validation Analysis

Time-based cross-validation was implemented to test model robustness. Performance metrics across folds showed:

- Consistent error values
- No significant performance degradation over time
- Reduced risk of overfitting

This confirmed that the refined model generalizes well to unseen future data.

7. Final Model Selection

Based on comparative evaluation, diagnostic analysis, and cross-validation results, the Histogram-based Gradient Boosting Regressor was selected as the final model. It demonstrated the best balance between accuracy, stability, and interpretability.

8. Conclusion

The model evaluation and refinement process significantly improved traffic prediction accuracy. Initial model weaknesses, particularly during peak hours, were addressed through feature enhancement and algorithm refinement. The iterative approach ensured that the final model is robust, reliable, and suitable for real-world traffic congestion forecasting.