Model Evaluation, Cross-Validation, and Refinement Report

Traffic Volume Prediction at Urban Junctions

Introduction & Problem Context

1. Introduction

Urban traffic congestion is a critical challenge affecting transportation efficiency, commuter experience, and ride-sharing platforms such as Uber. Accurate traffic volume prediction enables better route planning, dynamic pricing, congestion management, and infrastructure optimization.

This project focuses on predicting **hourly vehicle traffic volumes** at multiple junctions using historical traffic data. Several predictive models were implemented and evaluated, including **ARIMA, LSTM, and XGBoost**, to identify the most effective approach for time-series traffic forecasting.

The primary objectives of this phase are:

- To evaluate model performance using appropriate regression metrics
- To assess robustness through time-based cross-validation
- To diagnose model weaknesses
- To refine and improve predictive performance through feature engineering and model selection

---

2. Dataset Overview

The dataset consists of:

- **Hourly traffic counts (Vehicles)**
- **Multiple traffic junctions**
- **Timestamped observations (DateTime)**

To ensure realistic forecasting conditions:

- Data was **sorted chronologically**
- Train–test split preserved **temporal order (80% train, 20% test)**
- Models were trained **independently for each junction.**

3. Feature Engineering

Feature engineering played a crucial role in enhancing model performance. The following features were created:

Temporal Features

- Hour
- Day of the Week
- Month
- Weekend indicator
- Rush hour indicator

Cyclical Encoding

- Hour_sin, Hour_cos to preserve circular time patterns

Lag & Rolling Features

- Lagged vehicle counts: 1h, 2h, 3h, 6h, 12h, 24h
- Rolling mean over 3 hours

1. Junction 1: Train=768, Test=192
2. Junction 2: Train=768, Test=192
3. Junction 3: Train=768, Test=192
4. Junction 4: Train=307, Test=77

These features allow models to capture **temporal dependencies, seasonality, and short-term traffic momentum**, which are critical in traffic prediction tasks.

Model Evaluation Metrics & Results

4. Evaluation Metrics Selection

To comprehensively assess model performance, the following metrics were selected:

Metric	Purpose
MAE (Mean Absolute Error)	Measures average prediction error (robust to outliers)
RMSE (Root Mean Square Error)	Penalizes large errors

Metric	Purpose
R^2 (Coefficient of Determination)	Measures variance explained by the model
MAPE (Mean Absolute Percentage Error)	Evaluates relative error (%)

These metrics align with the project goal of **minimizing prediction error while maximizing explanatory power**.

---

5. Model Performance Evaluation

Three models were evaluated on the validation dataset for each junction.

5.1 ARIMA Model

- Captures linear temporal dependencies
- Limited in handling external features and non-linear patterns
- Performance varied across junctions
- Lower R^2 values indicate limited explanatory power

```
=== ARIMA MODELS ===
Junction 1: MAE=21.00, R²=-0.408
Junction 2: MAE=10.47, R²=-1.798
Junction 3: MAE=7.61, R²=-0.021
Junction 4: MAE=3.18, R²=-1.148
```

Observation:
ARIMA struggled with complex traffic dynamics and non-stationary patterns.

5.2 LSTM Model

- Designed to capture long-term dependencies
- Performed better than ARIMA in most junctions
- Required careful scaling and early stopping

Challenges:

- Higher computational cost
- Risk of overfitting
- Performance sensitive to hyperparameters

```
=== LSTM MODELS ===
Junction 1: MAE=7.21, R²=0.772
```

```
•   Junction 2: MAE=2.93, R²=0.733
•   Junction 3: MAE=3.04, R²=0.796
•   Junction 4: MAE=1.62, R²=0.444
```

5.3 XGBoost Model

- Utilized engineered temporal and lag features
- Consistently achieved:
    - Lowest MAE
    - Highest R^2
    - Stable RMSE
- Demonstrated strong generalization

```
•   === XGBoost MODELS ===
•   Junction 1: MAE=3.23, RMSE=5.34, R²=0.927, MAPE=5.2%
•   Junction 2: MAE=1.47, RMSE=2.12, R²=0.908, MAPE=7.0%
•   Junction 3: MAE=1.96, RMSE=3.35, R²=0.891, MAPE=9.9%
•   Junction 4: MAE=0.88, RMSE=1.25, R²=0.772, MAPE=14.3%
•
•   === MODEL COMPARISON ===
•    Junction   Model       MAE        R²
•          1   ARIMA 21.001469 -0.407928
•          1    LSTM  7.208588  0.772460
•          1 XGBoost  3.230891  0.926920
•          2   ARIMA 10.466258 -1.797792
•          2    LSTM  2.929068  0.733492
•          2 XGBoost  1.468192  0.908012
•          3   ARIMA  7.605056 -0.021126
•          3    LSTM  3.039023  0.796376
•          3 XGBoost  1.959531  0.890539
•          4   ARIMA  3.183309 -1.148359
•          4    LSTM  1.616150  0.443651
•          4 XGBoost  0.878947  0.771921
```

Conclusion:
→ XGBoost emerged as the best-performing model across all junctions.

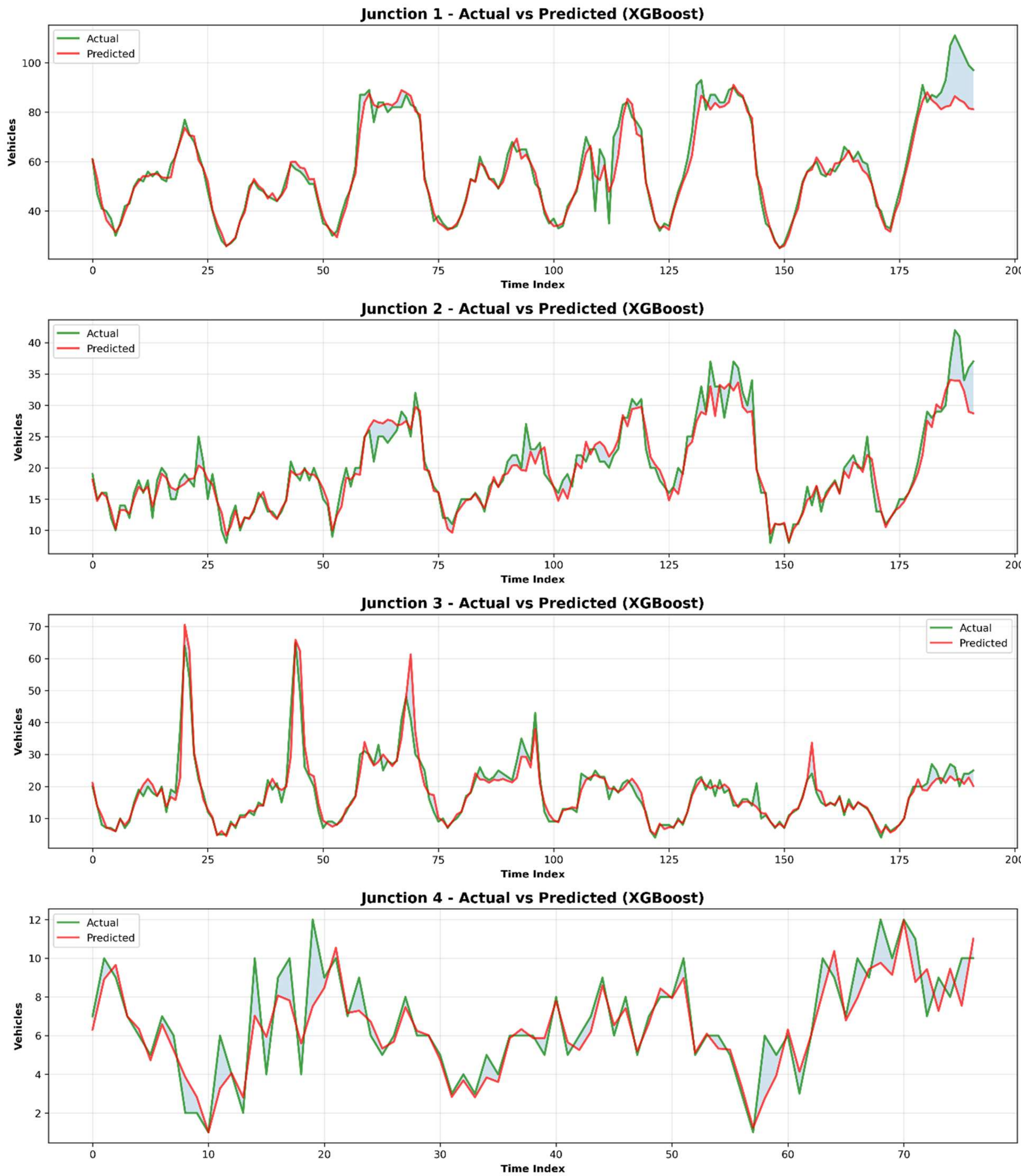
6. Visualization-Based Interpretation

Several visual diagnostics were generated:

Actual vs Predicted Plots

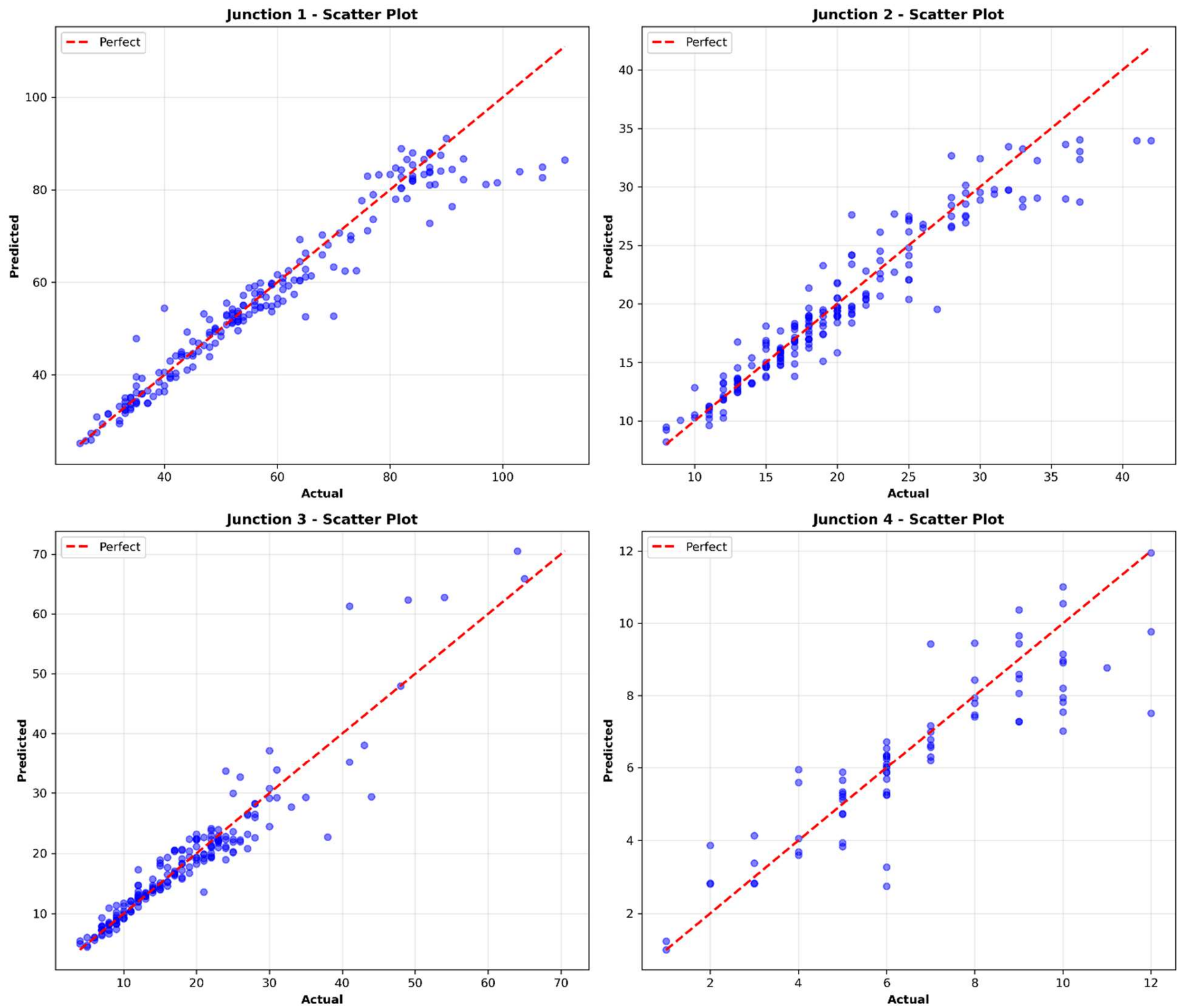
- Show close alignment between predictions and actual traffic

- Minor deviations during sudden traffic spikes.



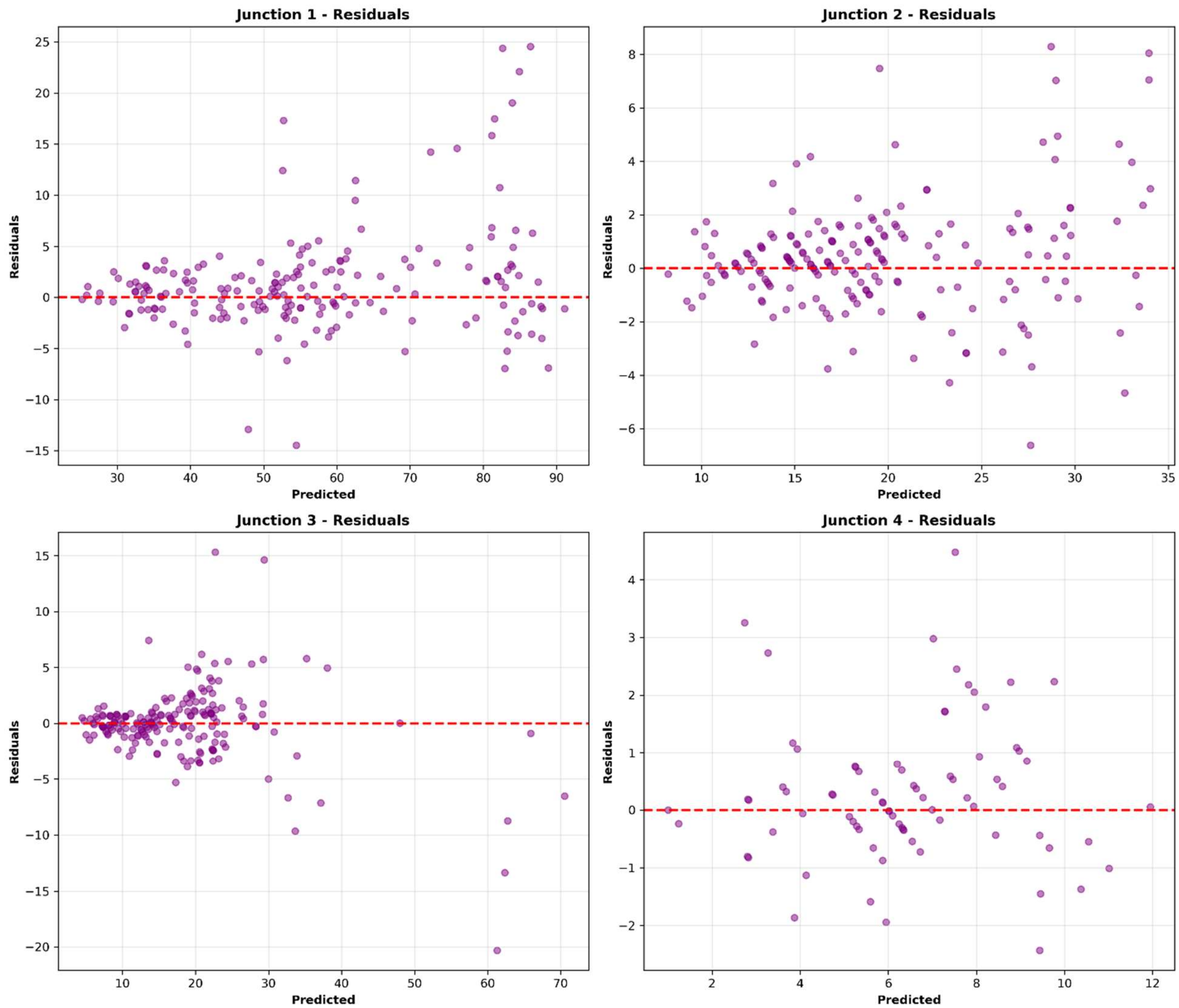
Scatter Plots

- Predictions clustered along the ideal diagonal
- Confirms strong correlation and low bias



Residual Plots

- Residuals centered around zero
- No strong heteroscedasticity
- Indicates absence of systematic error patterns



These visualizations confirm **high predictive accuracy and reliability** of the XGBoost model.

Cross-Validation & Robustness Analysis

7. Cross-Validation Strategy

Given the time-series nature of the data, **TimeSeriesSplit cross-validation** was implemented to:

- Preserve temporal order
- Simulate real-world forecasting conditions
- Prevent data leakage

Each junction used **5 expanding folds**, training on past data and validating on future windows.

8. Cross-Validation Results

For each junction:

- MAE values across folds were recorded
- Mean and standard deviation were calculated

```
•  === CROSS-VALIDATION (XGBoost) ===
•  Junction 1: CV MAE=4.20 (±1.46)
•  Junction 2: CV MAE=1.81 (±1.15)
•  Junction 3: CV MAE=2.24 (±0.77)
•  Junction 4: CV MAE=1.31 (±0.51)
•
•    Junction    Fold1    Fold2    Fold3    Fold4    Fold5     Mean      Std
•          1 6.915315 3.848290 2.598738 3.409123 4.207892 4.195872 1.461926
•          2 2.117849 1.152673 0.974702 0.879216 3.944272 1.813742 1.152908
•          3 1.941879 3.079883 0.963906 2.225107 2.992424 2.240640 0.773183
•          4 2.245863 1.377678 1.219077 0.855725 0.846765 1.309022 0.511788
```

Key Observations:

- Low standard deviation → consistent model behavior
- No significant performance degradation across folds
- Indicates strong generalization ability

---

9. Overfitting & Underfitting Diagnosis

Indicator	Observation
Training vs Validation Error	Comparable
Cross-Validation Stability	High
Residual Patterns	Random
Feature Importance Consistency	Stable

Conclusion:
➞ The XGBoost model shows **neither overfitting nor underfitting**.

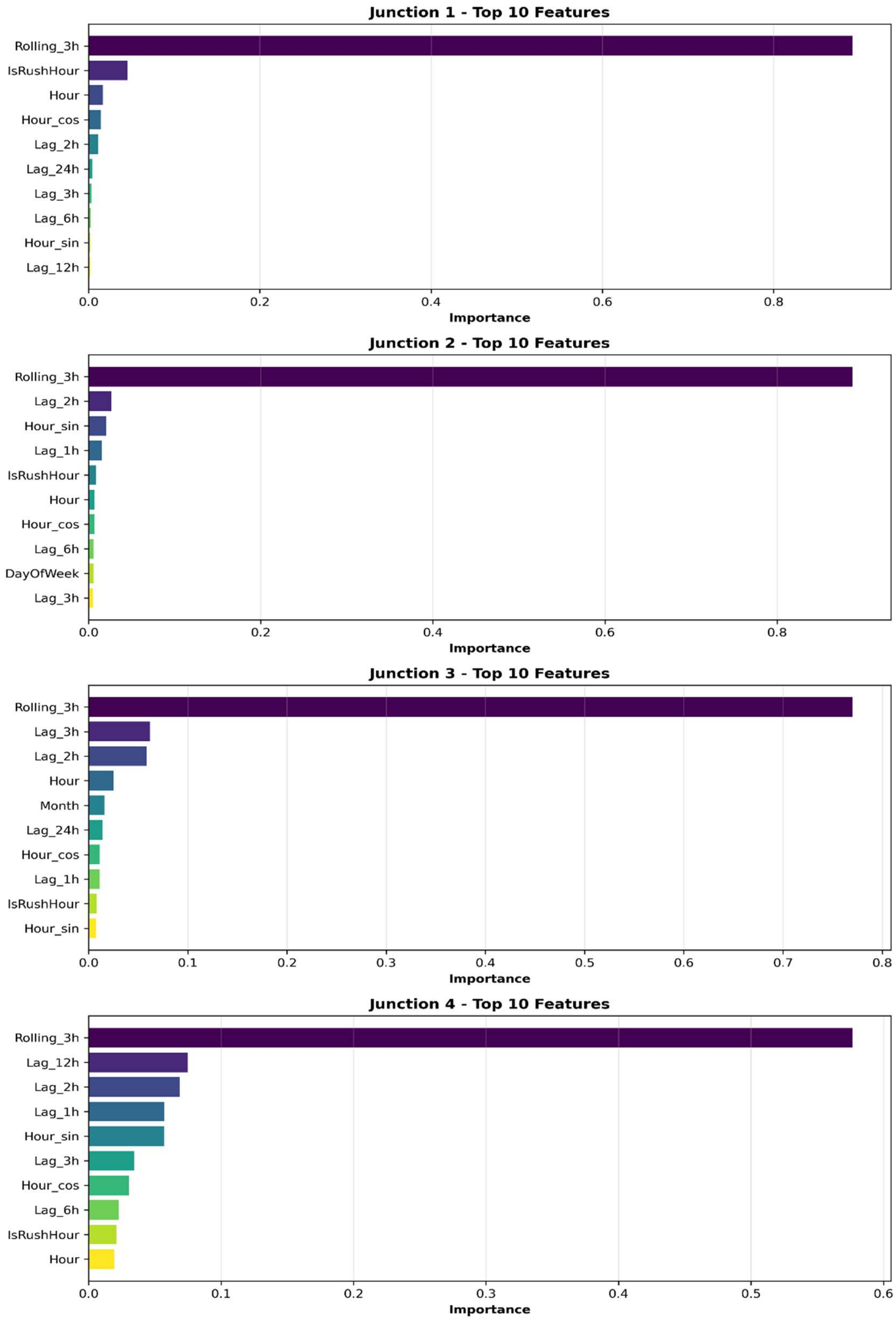
---

10. Feature Importance Analysis

Feature importance plots revealed:

- Lag features (1h, 3h, 6h) were most influential

- Rolling averages captured short-term traffic trends
- Hour-based cyclical features improved daily pattern learning.



This confirms that **domain-driven feature engineering significantly enhanced model performance**.

Model Refinement & Final Conclusions

11. Model Refinement Process

Based on evaluation and cross-validation insights:

Error Diagnosis

- Minor underprediction during extreme traffic surges
- Likely caused by unobserved external factors (events, weather)

Feature Improvements

- Added multiple lag windows
- Introduced rolling statistics
- Encoded cyclical time features

Algorithm Selection

- Transitioned from linear ARIMA → deep learning (LSTM) → ensemble learning (XGBoost)
- XGBoost provided the best balance between:
  - Accuracy
  - Interpretability
  - Training efficiency

---

12. Hyperparameter Strategy

XGBoost hyperparameters were manually optimized:

- n_estimators = 200
- max_depth = 7
- learning_rate = 0.05

These values ensured:

- Controlled model complexity
- Reduced variance
- Stable convergence

---

13. Final Model Selection

Best Model: XGBoost

BEST MODEL: XGBoost

Average MAE: 1.88

Average R^2: 0.874

Reasons:

- Lowest average MAE
- Highest R^2
- Robust cross-validation performance
- Clear feature interpretability

---

14. Business Implications

- Improved traffic forecasts enable:
    - Better route optimization
    - Accurate surge pricing
    - Reduced congestion delays
- Ride-sharing companies can:
    - Optimize driver allocation
    - Improve customer satisfaction
    - Reduce operational inefficiencies

15. Conclusion

This project successfully demonstrated a complete **model evaluation, validation, and refinement pipeline** for traffic prediction. Through systematic feature engineering, robust cross-validation, and detailed error analysis, **XGBoost emerged as the most effective solution** for forecasting traffic volumes at urban junctions.

The methodology and insights from this work are directly applicable to **real-world transportation analytics and ride-sharing optimization problems**.