Traffic Volume Prediction: Model Evaluation Report

Objective

To develop accurate and reliable predictive models for forecasting hourly traffic volumes at various road junctions using historical traffic and weather data.

Models Evaluated

ARIMA (statsmodels): Classical time series model.

XGBoost (xgboost): Tree-based ensemble model.

LSTM (sklearn-compatible): Deep learning for sequence modeling.

Gradient Boosting (sklearn): Interpretable ensemble method.

Final Model Comparison

ARIMA (Best): MAE = 8.60

ARIMA (Full): MAE = 9.82, MSE = 160.59, R^2 = -0.58

XGBoost: MAE = 1.24, MSE = 3.80, R^2 = 0.96

LSTM: MAE = 2.03, MSE = 9.79, R^2 = 0.90

Gradient Boosting: MAE = 1.97, MSE = 8.36, R^2 = 0.92

XGBoost outperformed all models with the lowest MAE and highest R^2.

Cross-Validation Summary

Gradient Boosting (5-fold CV):

MAE: 2.41 | RMSE: 2.89 | R^2: 0.92

XGBoost (5-fold CV):

Traffic Volume Prediction: Model Evaluation Report

Fold 1: MAE = 1.23, RMSE = 1.86, R^2 = 0.92

Fold 2: MAE = 1.46, RMSE = 2.74, R^2 = 0.88

Fold 3: MAE = 2.30, RMSE = 3.27, R^2 = 0.86

Fold 4: MAE = 1.82, RMSE = 3.18, R^2 = 0.87

Fold 5: MAE = 1.58, RMSE = 2.31, R^2 = 0.95

Average: MAE = 1.68 | RMSE = 2.67 | R^2 = 0.90

Cross-validation shows strong consistency and generalization.

Hyperparameter Tuning

Best Parameters:

- learning_rate: 0.1
- n_estimators: 100
- max_depth: 3
- subsample: 1.0

These settings achieved the best trade-off between bias and variance.

Key Takeaways

XGBoost is the best-performing model.

ARIMA underperformed on this multivariate, non-stationary data.

LSTM showed promise but adds complexity.

Feature engineering and hyperparameter tuning were key.

Final Recommendation

Traffic Volume Prediction: Model Evaluation Report

Deploy the XGBoost model with the selected hyperparameters for real-time traffic forecasting. Retrain periodically with updated data to maintain performance.