🚦 Forecasting Hourly Traffic Volumes Using XGBoost

Submitted by: Taranjeet Kaur

Menternship Title: UpGrad Campus Data Science Menternship

1. Executive Summary

This project aims to develop a predictive model to forecast hourly traffic volumes at road junctions using historical traffic data. By leveraging feature engineering and XGBoost regression, the model achieves moderate accuracy with an R^2 score of 0.4791. The final model can assist urban planners and traffic authorities in anticipating congestion and optimizing traffic flow.

2. Problem Statement

Accurate traffic volume prediction is crucial for managing urban mobility, reducing congestion, and improving road safety. The goal is to build a model that forecasts hourly traffic volume using historical data and contextual features.

3. Data Overview

- **Dataset:** Final_Integrated_Dataset.csv
- **Key Features:**
    - Vehicles: Hourly traffic volume
    - IsEvent: Event type (e.g., Holiday)
    - Time: Timestamp of observation
    - Weather-related variables (not used in final model)
- **Preprocessing:**
    - Removed missing values
    - Scaled target variable using MinMaxScaler

4. Feature Engineering

To enhance predictive power, the following features were engineered:

- Vehicles_lag1: Traffic volume from the previous hour
- Vehicles_ma3: 3-hour moving average of traffic volume
- IsHoliday: Binary flag for holidays

These features capture temporal patterns and event-based disruptions.

5. Model Development

- **Model Used:** XGBoost Regressor
- **Parameters:**
    - n_estimators = 100

    - max_depth = 5
    - learning_rate = 0.1
- **Split:** 80/20 train-validation split (no shuffle)

6. Model Evaluation

Metric	Value
MAE	0.0804
RMSE	0.1083
R^2	0.4791

The model explains ~48% of the variance in traffic volume. While not perfect, it provides a solid baseline for forecasting.

7. Error Analysis

- **Peak Hours:** Slight underprediction during high traffic periods
- **Holidays:** Increased error due to unpredictable spikes
- **Weather:** Not included, but may explain residual variance

8. Conclusion

The XGBoost model demonstrates moderate success in predicting hourly traffic volumes. Feature engineering played a key role in improving accuracy. Future improvements could include:

- Hyperparameter tuning
- Incorporating weather and road event data
- Using time-series models like LSTM