Predictive Modeling for Traffic Volume Forecasting

Using Machine Learning (XGBoost) on Real-World Traffic Data

Introduction:

- Traffic congestion is a critical urban challenge.
- Accurate prediction of traffic volume helps in better road planning and congestion management.
- This project builds a machine learning model to predict traffic volume using historical and contextual data (like time, weather, incidents).
- The XGBoost model is used for robust regression performance.

Dataset Overview:

- **Dataset columns:**{

Timestamp, location, traffic_volume, temperature, weather, is_holiday, incident, event}

- **Target Variable**:{

traffic_volume }

- Features include both numerical and categorical data.
- Preprocessing was done to extract useful features from the timestamp and other variables.

Feature Engineering:

- **Extracted time-based features:**

  {hour, day_of_week}

- **Created lag features:**

  {lag_1, lag_2, lag_3}

- **Created rolling mean features:**

  {rolling_mean_3, rolling_mean_6}

- **Encoded categorical variables:**

  {weather, event using one-hot encoding}

Data Preprocessing:

- Checked for missing values and cleaned the data.
- Converted timestamp to datetime and extracted useful components.
- Applied one-hot encoding for categorical features.
- Prepared the final feature set for modeling.

Model Training:

- Model used: XGBoost Regressor
- Parameters: n_estimators=100, learning_rate=0.1
- Split data: 80% training, 20% testing
- Target: traffic_volume

Model Evaluation:

- ✓ Evaluation Metrics:
  - ❖ MAE (Mean Absolute Error)
  - ❖ RMSE (Root Mean Squared Error)
  - ❖ R^2 Score
- ✓ Predicted traffic volume vs actual shown using line plot.
- ✓ Visualized performance over first 100 samples.

Feature Importance:

- XGBoost provides feature importance scores.
- Most important features were:
- hour, day_of_week
- recent lag values and rolling_mean
- contextual info like weather and incident
- Bar chart used to display importance scores.

Conclusion & Future Work

- XGBoost model successfully predicts traffic volume with good accuracy.
- Useful for real-time traffic monitoring and planning.

    - Future enhancements:
        - Include real-time sensor data
        - Use deep learning (e.g., LSTM)
        - Deploy as a live dashboard or API

THANK YOU