Project: Traffic Incident Severity Prediction

Objective

Develop a supervised machine learning model that predicts the **severity level (1–4)** of a traffic incident based on its spatial, temporal, and contextual features. This aids dispatchers and first responders in triaging incidents and allocating resources efficiently.

1. Data Preparation

1. **Load and clean data**
    - Parse `Start_Time`/`End_Time` as datetime.
    - Drop or impute missing geographic coordinates.
    - Remove duplicate records.
2. **Feature engineering**
    - **Temporal features**:
        - Hour of day (`Start_Time.dt.hour`)
        - Day of week (`Start_Time.dt.weekday`)
        - Incident duration in minutes: (End_Time − Start_Time). total_seconds()/60
    - **Spatial features**:
        - Start latitude/longitude directly or binned grid cells
        - Distance impacted (`Distance(mi)`)
    - **Traffic-control flags**: use boolean columns (`Traffic_Signal`, `Roundabout`, etc.)
    - **Light condition**: encode `Sunrise_Sunset` and twilight phases as categorical (one-hot)
    - **Text features (optional)**:
        - Encode `Description` via TF-IDF or keyword presence (e.g., “lane blocked,” “accident on”)
3. **Train/test split**
    - Stratify by `Severity` to ensure all levels appear in both sets.

- Typical 70/30 or 80/20 split.

2. Model Selection

1. **Baseline models**
   - Logistic regression (multinomial) or decision tree as a simple benchmark.
2. **Advanced models**
   - Gradient-boosted trees for strong performance on tabular data.
   - Random forest classifier as an ensemble alternative.
3. **Hyperparameter tuning**
   - Use cross-validation with grid search or Bayesian optimization to tune tree depth, learning rate, number of estimators, etc.

3. Evaluation Metrics

- **Accuracy**, **F1-score** (macro) for overall performance across all severity levels.
- **Confusion matrix** to diagnose which severity levels are often misclassified.
- **ROC AUC** per class via one-vs-rest for additional insight.

4. Deployment & Interface

- **Web dashboard** (e.g., Streamlit or Flask) where users:
  - Upload or enter incident details (time, location, traffic features).
  - Receive a predicted severity level and confidence scores.
  - Visualize feature-importance via SHAP to explain model decisions.

This project provides actionable severity predictions and interpretable insights to enhance traffic incident response and public safety.