Remaining Objectives

High-Risk Combinations (Final Phase)

1. Use PCA to reduce dimensionality and visualize clusters of high-risk combinations (profession + department).
2. Run K-Means clustering on encoded categorical data to find hidden high-risk groupings.

Predict Severity of Incidents

3. Feature Engineering
    - Convert categorical variables (e.g., profession, department, perpetrator type, violence type) to numerical using OneHot or Label Encoding.
    - Handle time-based features (e.g., extract hour, weekday, or month from `event_time`).
    - Optionally, use text features (e.g., response_action) via keyword flags.
4. Model Training
    - Train classification models: Logistic Regression, Random Forest, SVM.
    - Use cross-validation or train/test split.
5. Evaluation
    - Compute confusion matrix, accuracy, F1-score, AUC.
    - Visualize results (confusion matrix heatmap, ROC curve, etc.).
6. Feature Importance
    - Analyze which features contribute most to predicting severity.
    - Visualize importance (bar chart or SHAP if applicable).

Team

Ruobing Wang

Yiqing Ma

Yingxin Song