### **Crime Hotspot Detection and Prediction Report**

#### **1. Overview**

This analysis aims to identify crime hotspots in a given geographical region using crime data. A crime hotspot is defined as an area with a high frequency of criminal activity, which can be used for better resource allocation and crime prevention. In this report, we utilize a combination of geospatial analysis and machine learning to predict crime hotspots using a dataset of crime locations with their corresponding latitude and longitude.

#### **2. Data Preparation**

* **GeoDataFrame Creation**:
* The input crime data, which includes latitude and longitude coordinates for each crime incident, was converted into a GeoDataFrame using the geopandas library. This allowed us to represent the data spatially, enabling spatial analysis and visualization of the crime incidents on a map.

* **Bounding Box Calculation**:

A bounding box that encompasses all the crime data points was calculated dynamically using the total\_bounds attribute of the GeoDataFrame. This bounding box helps define the extent of the grid where we will generate spatial cells for further analysis

#### **3. Grid Creation**

A grid was created over the bounding box of the crime data to divide the area into smaller cells. Each cell represents a geographical unit within which the crimes will be counted. The grid cell size was defined as 0.002 degrees (approximately 250x250 ft), which was chosen to balance the granularity of the analysis and computational efficiency.

* **Spatial Join with Crime Data**:

The grid was spatially joined with the crime data to determine which crime incidents fall within each grid cell. This join allowed us to count the number of crimes in each grid cell, which is essential for identifying crime hotspots.

#### **4. Crime Hotspot Identification**

* **Hotspot Calculation**:

To classify grid cells as "hotspots," a threshold based on the 95th percentile of crime counts across the grid was used. Cells with a crime count greater than or equal to this threshold were designated as hotspots.

* **Visualization of Hotspots**:

The crime hotspots were visualized on a map using matplotlib and geopandas. The map displayed the grid cells with the crime hotspots highlighted.

python

Copy code

grid\_gdf.plot(ax=ax, column='hotspot', cmap='coolwarm', legend=True, alpha=0.6)

#### **5. Machine Learning Model for Prediction**

* **Feature Engineering**:

Two features were created for machine learning classification: crime\_count (number of crimes in each grid cell) and neighbor\_crime\_count (the sum of crimes in neighboring grid cells). The idea was to capture both the local crime count and the influence of surrounding areas in determining whether a grid cell is a hotspot.

* **Model Training**:

A RandomForestClassifier was used to predict crime hotspots based on the two features. The dataset was split into training and testing sets using an 80-20 split, and the model was trained on the training set.

* **Model Evaluation**:

After training, the model was evaluated on the test set using accuracy and a classification report to measure its performance. The model achieved a certain level of accuracy in predicting hotspot areas.

#### **6. Visualization of Predicted Hotspots**

* The predicted hotspots were visualized on the map, with the predicted hotspots highlighted. This provided a clear view of where the model expects high crime areas to be.