**Project Proposal**

**Detection of Patterns and Trends in Crime Data in Greater Toronto Area Using Unsupervised Learning**

****

**Overview**

The detection of patterns and trends in crime data is critical for understanding the evolving landscape of public safety. Our project aims to identify and analyze changes in crime patterns in the Greater Toronto Area (GTA) using unsupervised learning techniques. By leveraging historical crime data, we will apply clustering algorithms to detect hotspots, analyze trends over time, and uncover hidden patterns that may be linked to external factors such as time of day, location, or recurring events. The goal is to provide valuable insights into how crime behaviors fluctuate and identify potential factors driving these changes, enabling more informed decision-making for law enforcement and policymakers in the GTA.

**Objectives**

* **Analyze Crime Patterns:** Identify trends and patterns in various types of crime in the Greater Toronto Area (GTA), using machine learning to uncover hidden insights.

**Tools & Technologies**

* **Machine Learning:**
  + **Scikit-learn** for applying clustering algorithms (e.g., K-Means, DBSCAN, Random forest) to detect crime hotspots and trends.
* **Data Manipulation:**
  + **Python Pandas** for data cleaning, preprocessing, and manipulation.
* **Visualization Libraries:**
  + **Matplotlib** (Python) for visualizing temporal crime trends and insights.

**Data Source**

* **Toronto Police Service Crime Data:**
  + Historical crime records with more than 20000 data points, containing spatial (GeoJSON) and temporal features. The data covers multiple years, enabling the analysis of crime patterns in 2021, 2022 and 2023.

**Expected Outcomes**

* **Pattern Recognition:** Detection of crime trends and hotspots in the GTA, based on spatial and temporal features.
* **Predictive Insights:** Machine learning models that can predict future crime trends, enabling data-driven decision-making for law enforcement.
* **Visual Insights:** Geospatial visualizations showing how crime hotspots and patterns have evolved over time.

**Expected Deliverables**

* **Machine Learning Analysis:** Clustering and pattern detection using Scikit-learn to uncover trends in crime data.
* **Interactive Visualizations:**
  + At least three different visual views, including:
    - **Geospatial maps** of crime hotspots using Leaflet.
    - **Temporal visualizations** of crime trends using Matplotlib.
    - **Interactive graphs** for user-driven exploration of crime data with Plotly.
* **README.md:** A detailed document with an overview of the project, instructions for use, ethical considerations, and references.

**Conclusion**

This project will provide valuable insights into crime patterns and trends in the Greater Toronto Area, applying unsupervised learning to identify hotspots and hidden patterns. By leveraging machine learning and advanced data visualization techniques, it will deliver actionable insights that could inform law enforcement strategies and public safety planning, ensuring a data-driven approach to crime prevention.

|  |
| --- |
| **Data Sources:**  <https://data.torontopolice.on.ca/pages/open-data>  API base url: <https://services.arcgis.com/S9th0jAJ7bqgIRjw/ArcGIS/rest/services/Major_Crime_Indicators_Open_Data/FeatureServer/0/query> |
| **Team Members**  Neal L, Lavanya B, Mahnaz K, Julian P, Rasadari R |
|  |