**Project Proposal: Crime Analysis and Prediction Framework**

**1. Problem Description:**

We propose to address the problem of rising crime rates and inefficient crime management strategies in urban areas, with a focus on Boston. The primary goal is to develop a comprehensive framework that utilizes machine learning and natural language processing techniques to predict crime rates, recognize crime patterns, classify crime types, identify crime hotspots, and analyze community sentiment towards crime and safety.

**Inputs:** Historical crime data, offense codes, district information, temporal data, and social media data.

**Outputs:** Predicted crime rates, crime type classifications, identified crime hotspots, sentiment analysis outcomes.

**Data Used:** The project will utilize two main datasets: **crime.csv** for historical crime records and **offense\_codes.csv** for crime classification codes specific to Boston.

**Significance:** This project is significant as it aims to assist law enforcement agencies in making data-driven decisions, improve public safety, and foster a better understanding of community perceptions regarding crime.

**2. Algorithms:**

1. **Crime Rate Prediction Model:** We will use **Support Vector Machines (SVM)** and **Random Forest** algorithms for predicting crime rates. SVM is chosen for its effectiveness in handling high dimensional data and its ability to model non-linear decision boundaries. Random Forest is selected for its robustness, ability to handle unbalanced data, and feature importance evaluation.
2. **Crime Pattern Recognition and Classification:** For this, we will employ **unsupervised learning algorithms** such as hierarchical clustering to identify underlying patterns and classify different types of crimes. This method helps in understanding the nature of crimes without predefined labels.
3. **Crime Hotspot Detection and Analysis:** We will use **DBSCAN,** and **K-Means clustering** to identify and analyze crime hotspots. DBSCAN is appropriate for identifying clusters of varying shapes and sizes, while K-Means is efficient for large datasets.
4. **Sentiment Analysis of Community Perception on Crime:** We will utilize **Natural Language Processing (NLP)** techniques to analyze public sentiment towards crime and safety using social media data. This will involve text mining, sentiment analysis, and topic modeling to understand public opinions and concerns.

Note: We have narrowed it down to this Four different algorithms that can be applied to the dataset. We aim to finish at least two out of this four.

**Previous Work:** Similar algorithms have been used in various studies for crime rate prediction, pattern recognition, hotspot detection, and sentiment analysis. Our approach is unique as we integrate these methodologies into a unified framework tailored for Boston's urban landscape.

**3. Expected Results:**

We anticipate demonstrating the effectiveness of our framework by accurately predicting crime rates, identifying and classifying crime patterns, detecting crime hotspots, and analyzing community sentiment. We will compare our results with existing crime statistics and studies to validate our model's accuracy and efficiency.

**Comparisons and Risks:** We will compare our results with traditional crime analysis methods and check for improvements in prediction accuracy and resource allocation. Risks include potential inaccuracies in data and unforeseen challenges in algorithm implementation. To mitigate these risks, we will continuously refine our models and seek expert guidance.

**Team Contributions:**

* Kaumudi Rawal will focus on the Crime Rate Prediction Model using SVM and Random Forest.
* Priyank Patel will handle Crime Pattern Recognition and Classification using hierarchical clustering.
* Once two algorithms are implemented accurately, our ambitious goal is to implement the further mentioned algorithms.

**References:** [

https://www.kaggle.com/datasets/AnalyzeBoston/crimes-inboston/data]