A crime is a deliberate act that can cause physical or psychological harm, as well as property damage or loss, and can lead to punishment by a state or other authority according to the severity of the crime. To be better prepared to respond to criminal activity, it is important to understand patterns in crime. In this project, we analyze crime data from the Major City, data scraped from publicly available records & police websites. At the outset, the task is to predict which category of crime is most likely to occur given a time and place in the city. The use of AI and machine learning to detect crime via sound or cameras currently exists, is proven to work, and expected to continue to expand.

The use of AI/ML in predicting crimes or an individual’s likelihood for committing a crime has promise but is still more of an unknown. The Biggest challenge will probably be “proving” to politicians that it works. When a system is designed to stop something from happening, it is difficult to prove the negative. Companies that are directly involved in providing governments with AI tools to monitor areas or predict crime will likely benefit from a positive feedback loop. Improvements in crime prevention technology will likely spur increased total spending on this technology. In this Project it attempts to make our classification task more meaningful by Merging Multiple classes into larger classes. Finally, we report and reflect on our results with different classifiers, and dwell on avenues for Future work.

Crime Data Mining is a project made for public safety and protection related to crime, and a better understanding of crime is beneficial in multiple ways: it can lead to targeted and sensitive practices by law enforcement authorities to mitigate crime, and more concerted efforts by citizens and authorities to create healthy neighbourhood environments. With the advent of the Big Data era and the availability of fast, efficient algorithms for data analysis, understanding patterns in crime from data is an active and growing field of research. The increasing use of computerized systems to track crimes, computer data analysts have started helping the law enforcement officers and detectives to speed up the process of solving crimes.

The inputs to this algorithms are time (hour, day, month, and year), place (latitude and longitude), and class of crime :

**Act 379 – Robbery,**

**Act 13 – Gambling,**

**Act 279 - Accident,**

**Act 323 – Violence,**

**Act 302 – Murder,**

**Act 363 – Kidnapping.**

Output is the class of crime that is likely to have occurred. We try out multiple classification algorithms, such as KNN (K-Nearest Neighbors), Decision Trees, and Random Forests. Such clusters can be visually represented using a geo-spatial plot of the crime overlayed on the map of the police jurisdiction. The densely populated group of crimes is used to visually locate the ‘hot-spots’ of crime. We also perform multiple classification tasks – we first try to predict which of 6 classes of crimes are likely to have occurred with the help of Geo-spatial techniques and later try to differentiate between violent and non-violent crimes.

**Module Description :**

Used three different classification problems to solve, which we proceeded to attack with an assortment of classification algorithms.

The following are the algorithms which are used:

**Dataset & Preprocessing**

**KNN( K- Nearest neighbors)**

**Decision Tree**

**Random Forest**