**Predictive Analysis of Violent Crime in Baltimore City**

By Team “Data Miners”  
  
Bhavana Tadepalli - U89903796  
Raja Sreeharsha Meda-U24682304  
Praveen Kumar Lankala- U17584089  
Venkata Krishna Sri Chandana Manuri - U97875301  
Ramesh Nalla - U48307125

A map of baltimore neighborhoods

Description automatically generated

**Background:**

Baltimore, like many urban areas, grapples with a range of crime-related issues that impact the safety and well-being of its residents. From violent crimes to property crimes, the city faces ongoing challenges in maintaining law and order. The key problems include:

* A lack of comprehensive understanding of the root causes of crime.
* Insufficient knowledge about the potential crime locations.
* Inadequate awareness of the specific times when crimes occur.

**Business Proposal:**

To overcome the above problem, we use machine learning techniques to proactively forecast potential crime hotspots and the times of crime happening. Our analysis encompasses factors such as geographical patterns, and historical crime data. By leveraging various machine learning algorithms, we aim to enhance the accuracy of our predictions, ultimately assisting law enforcement and safety agencies in proactive crime prevention.

When trying to figure out when violent crimes are more likely to occur, we study past crime data to find patterns in the times when they usually happen. This helps police use their resources effectively by focusing on the times when crimes are most likely to occur. It gives them a way to plan and prevent crimes from happening. The data also helps them address the reasons people commit crimes, like helping the community and providing support services.

**Data source:** <https://data.baltimorecity.gov/datasets/baltimore::part-1-crime-data/explore?location=33.172431%2C-53.426424%2C5.88>

**Variables used:**

**CrimeDateTime:** The time at which crime occurred.

**Description:** The type of crime that occurred

**Inside\_Outside:** The place where the crime happened inside or outside.

**Gender:** The gender of the Perp

**Age:** The age of the perp

**Race:** The race of the perp

**Ethnicity:** The ethnicity of the perp

**New\_District:** District of the place

**Latitude:** The latitude of the location where the crime happened

**Longitude:** The longitude of the location where the crime happened

Here the variable ‘Description’ gives information regarding the type of crime that happened. Now we are considering only ‘SHOOTING’, ‘HOMICIDE’, ‘RAPE’,’ AGGRAVATED ASSAULT’, and ‘ARSON’ as violent crimes and converting whole data into a binary category. Where 1 is taken as a violent crime and 0 as the crime. The **Target** variable is violent\_crime.

.

**Data after Preprocessing:**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a table

Description automatically generated**

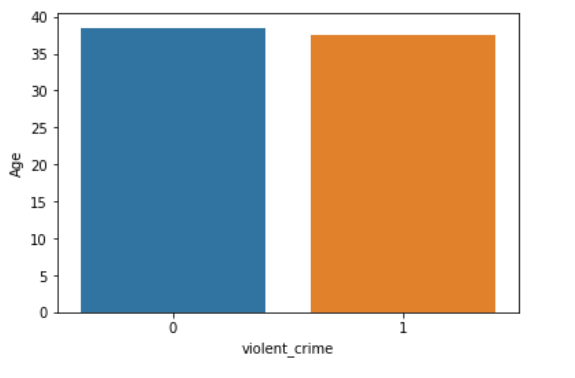
**We divided the data into 70:30. As time variables are included, didn’t split the data randomly. Arranging the data in chronological order by sorting the data by time\_hour, Day, Month, and Year. So first 70% of the data is considered as train data and the remaining as test data.**

**Methods:**

**Models Used:**

This predictive model is constructed by considering an array of crucial predictive variables. These factors encompass elements such as geographical location, socioeconomic indicators, and historical crime data. By leveraging a combination of machine learning algorithms, including Logistic Regression, KNN classifiers, Decision Tree Classifiers, Random Forest Classifiers, XGBoost Classifiers and AdaBoost Classifiers. Our goal is to establish a robust predictive model that can effectively gauge the probability of a crime taking place. This holistic approach harnesses the strengths of these algorithms to enhance the model's predictive accuracy, ensuring that law enforcement and public safety agencies can better anticipate and respond to potential criminal events.

**Results:**



**From this graph, we can conclude that age is not a determining factor in the occurrence of crime.**

A graph with blue and orange bars

Description automatically generated

**Based on the graph above, it can be observed that a significant majority of the crimes are committed by individuals who identify as Black or African American.**

A graph of blue and orange bars

Description automatically generated

**The graph above suggests that there is a higher occurrence of crimes in the northeastern and southeastern regions.**

A blue line graph with numbers

Description automatically generated

**The chart above illustrates the relationship between the number of crimes and the corresponding years.**

**The graph depicts the correlation between number of months and the number of crimes occurring.**A graph with a line

Description automatically generated

**The graph depicts the correlation between the number of hours and the frequency of crimes occurring.**

A line graph with numbers

Description automatically generated

**The graph shows how the time of hours relates to the occurrence of crimes.**

**Scatterplot of the Latitude and Longitude with Target**

A map of the land

Description automatically generated with medium confidence

**Discussion:**

**By the above results, we can say that XG boost is providing the best results.**

A screenshot of a graph

Description automatically generated

**Confusion Matrix:**

**A blue squares with numbers

Description automatically generated**A diagram of negative and false negative

Description automatically generated

**Feature Importance:**

According to the below values, we can say that category “outside” of the “Inside\_outside” variable has more impact on the model followed by category “M” in gender has the second most impact on the model as the year has the third most impact on the model. Month and day have less impact on the model.

**A screenshot of a table

Description automatically generated**

**A graph of a number of people

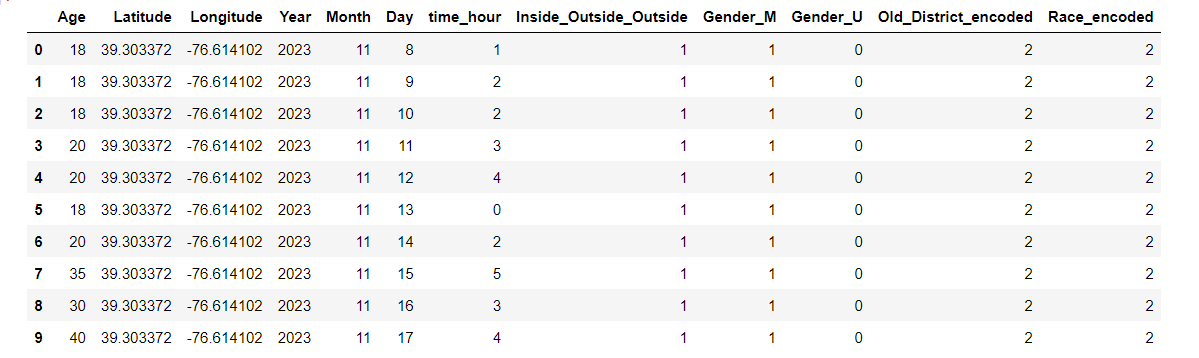
Description automatically generated with medium confidence**

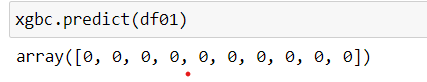
**Conclusion:**

In this study, we developed a machine-learning model to predict violent crime occurrences in Baltimore. Our Model achieved an accuracy of 0.75 with a precision of 0.545, a recall of 0.211, and an F1 score is 0.305. These metrics indicate that the model is performing well in identifying potential violent crime incidents. The analysis reveals that features such as Inside out, gender, and year have the most significant influence on this model.

**Forecasting:**

**Predicted the violent crimes occurrence for the next 10 days.**



****

**By this we can say that there are no violent crimes happening in these longitudinal and latitudinal areas (1127 St Paul St, Baltimore, MD 21202) for the next 10 days.**