Crimson Eye - A Data Driven Approach to Crime Analysis

Dr. Shamsad Parvin
CSE 587 - Data Intensive Computing

Tarun Reddi
bhanucha@buffalo.edu

Charvi Kusuma
charviku@buffalo.edu

University at Buffalo

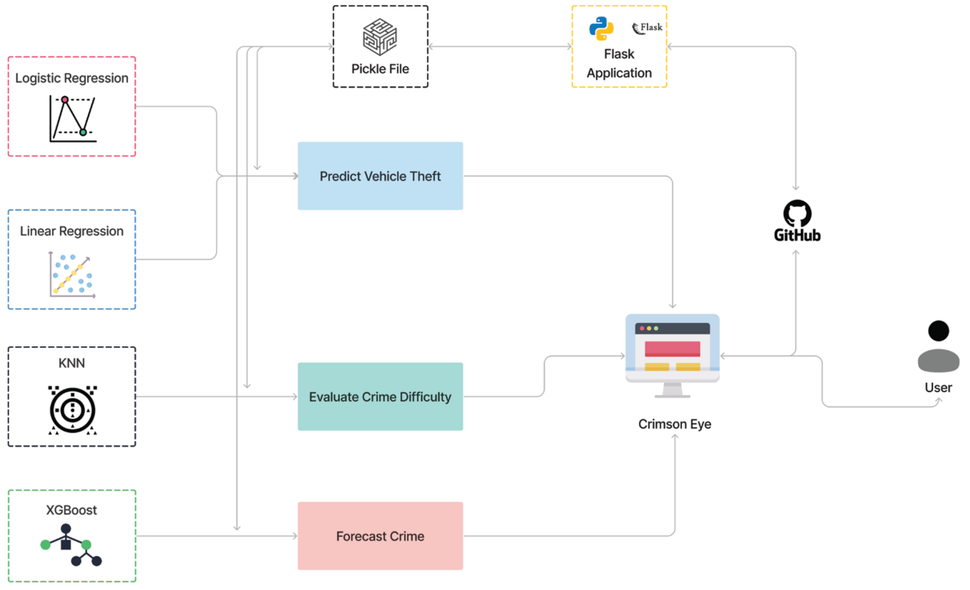
Overview

Strategic Decision Making: CrimsonEye supports users and police departments in making data-driven decisions about criminal activities, serving as a key tool for crime trend analysis.

Efficient Resource Management: The platform enhances the allocation of police resources through predictive analytics, streamlining responses and improving public safety.

Enhanced Community Awareness: CrimsonEye enables users to easily assess criminal risks in their areas, fostering community safety and vigilance with insights derived from historical crime data.

Data Source: City of Los Angeles Open Data



Process Sequence

Dataset Overview

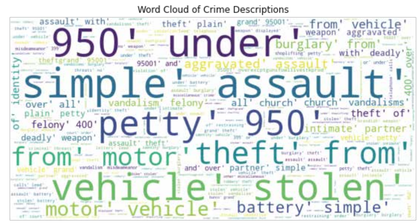
Title: Crime Data from 2020 to Present

Dimensions: 811663 Rows, 28 Columns

The data is sourced from original crime reports, emphasizing its real-world relevance. These reports are generated by the Los Angeles Police Department (LAPD) and are considered a reliable and authoritative source of information.

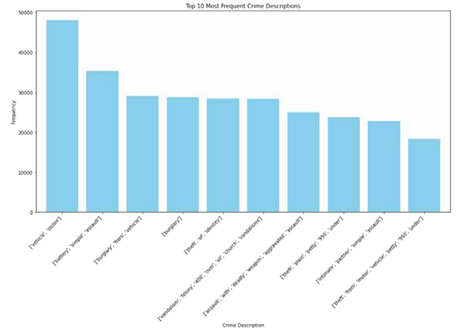
Short Insights

- Word Cloud of Crime Descriptions



The graph shows the top 10 most frequent crime descriptions in Los Angeles from 2020 to the present day. This information can be used to develop crime prevention strategies and to allocate resources to law enforcement and social services.

- Top 10 Most Frequent Crime Types



Word cloud highlights the most frequent crime descriptions in Los Angeles, revealing insights into prevalent crimes. Property crimes, particularly vehicle theft, burglary (including burglary from vehicles), and petty theft, dominate the dataset.

- **Logistic Regression**
  - In logistic regression, the probability ***P*(Y=1)** of an instance belonging to a particular class is modeled as a function of the independent variable ***x*** and model coefficients ***β0, β1***, using the logistic function.

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

  - **Use Case:** The vehicle Theft prediction feature in our application, enables us to identify the risk involved with our vehicle based on the area code the time you are parking, and the premises you are located.

- K-Nearest Neighbors
  - KNN operates by finding the 'k' closest data points (neighbors) to a given input point, based on a distance metric like Euclidean distance.
  - Euclidean Distance formula calculates the distance in an n-dimensional space, where qi and pi are the coordinates of the two points in each dimension.

$$\sum_{i=1}^{n} (q_i - p_i)^2$$

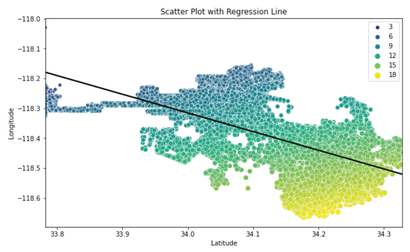
  - Use Case: Exploiting the capabilities of KNN, we predict the difficulty of a crime investigation based on past crime occurrences.

- **Linear Regression**
  - A statistical method for modeling the relationship between a dependent variable and one or more independent variables.
  - **Formula:**

$$(y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon)$$

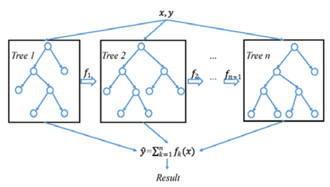
$$(y = \beta_0 + \beta_1 x + \varepsilon)$$

  - In the regression model, ***y*** is the dependent variable, ***x*** is the independent variable, ***β0*** is the intercept, ***β1*** is the slope coefficient, and **ε** is the error term. In case of multiple regression ***βi,*** will be the coefficients of independent variables ***xi,*** where ***i= 1,2,3, .. n***
  - **Use Case:** Our application uses this model to analyze the Area Code of Los Angeles based on the crime coordinates.



- **XGBoost**
  - XGBoost algorithm, which stands for Extreme Gradient Boosting, is complex and involves multiple mathematical concepts and formulas. At its core, XGBoost implements gradient-boosted decision trees, designed for speed and performance.
  - **Objective Function**: The objective function in XGBoost combines a loss function and a regularization term. It is given by:

$$\text{Obj}(\theta) = L(\theta) + \Omega(\theta)$$



User: People

Input: Location, Time, Premise Type, Area Code

Analytical Methods: Linear & Logistic

Outcome: Determining the risk of your vehicle being stolen

User can assess the likelihood of their vehicle being stolen based on where, when, and the type of place, using predictive models like Linear and Logistic Regression.

Predict Vehicle Theft

User: Police Force/ People

Input: Time, Premise Description, Area Code, Location Type

Analytical Methods: XGBoost

Outcome: Based on your area, your street location and the premise description, we aim to predict the most prevalent crime in that locality

Suppose a user is new to a location, he can be one step ahead by using our tool and aware of his location crimes.

Forecast Criminal Activity

User: Police Forces

Input: Time Occurred, Area Code, Area Name, Reporting District, Crime Code, Victim Age, Victim Sex, Victim Descent, Premise Type, Latitude, Longitude, Location Type.

Analytical Methods: K Nearest Neighbors

Outcome: Evaluate the current crime's complexity by predicting difficulty through an analysis of past instances with similar characteristics.

Estimate Crime Difficulty

Expanding the Application

- Dual-function application
- Assists LA police with crime rate management and neighborhood monitoring
- Equips LA residents with local crime data for increased safety awareness

Contribute to Public Safety - Data Collection

We have included an additional feature to acquire recent data. Since our models have been trained on the limited crime data from the year 2020, we would like to expand the dataset for the robustness of crime data analytics.

Adaptive User Interface

Resources