

Crime Rate Prediction in LA

Deepak Bhadouria, Ritu Kukreja, Surendra Pothuri

Introduction:

The number and forms of criminal activities are increasing at an alarming rate. It is a deliberate act that can cause physical or psychological harm, including property damage or loss in many cities, and LA is no exception. As a result, law enforcement and community organizations are constantly looking for ways to reduce crime and increase public safety. One promising approach is to use data analytics and machine learning models to predict crime rates and assist associated authorities in preventing them. By leveraging data-driven insights, predictive models can help identify areas or situations that are likely to see an increase in criminal activity. Additionally, these models can be used to provide a better understanding of the underlying causes of crime and to develop more effective strategies for preventing and responding to crime. Cloud computing can be used to store and analyze large volumes of data, while machine learning models can be used to identify patterns in that data and make more accurate predictions. Through this approach, law enforcement and community organizations will be better equipped to plan and deploy resources in order to reduce crime and improve public safety.

Sample Dataset:

We would work on the dataset available at kaggle. The data set includes information on the location, time, type of crimes etc, dating as far back as several years. This dataset will be used to analyze crime trends in the city, as well as to help inform policy decisions and strategies.

https://www.kaggle.com/datasets/hemil26/crime-in-los-angeles?select=crime_in_la.csv

EDA:

We will use EDA to explore the data, look for patterns, and understand the relationships between different variables. We will then use basic data analysis methods such as summary statistics, visualizations, and correlations to explore the data such as correlation between numerical features, visualization for categorical variables, standardization.

Purpose:

The purpose of our project is to get a thorough knowledge of PySpark's Data Processing, Analysis, and modeling tools such as Data Transformation, Data Visualization, Exploratory Data Analysis, and Feature Engineering including MLlib. We will be employing visualization techniques to explore any patterns, trends, or correlations in the data that may point to risk areas and times for criminal activity. Moreover, we would use EDA to explore the data and identify outliers that may indicate suspicious activity. Finally, we will work on machine learning algorithms that can be used to build informative models that can predict future criminal activity, allowing associated agencies to take preventative action and potentially reduce crime.