1. Introduction

The Washington Post conducted a comprehensive data collection effort focused on criminal homicides spanning the last ten years across 50 major American cities. This dataset encompassed details such as the incident location, whether an arrest was made, and, in most cases, basic demographic information about the victims. Reporters faced the task of organizing this data, which arrived in various formats including physical documents. They invested months in the process of cleaning and standardizing the information, cross-referencing homicide counts and closure rates with FBI data to ensure accuracy. In instances where law enforcement agencies provided only partial details on the homicides, reporters turned to public records such as death certificates, court documents, and medical examiner reports to fill in the gaps. These data have more details compared to the annual federal homicide data compiled by the FBI from police agencies across the country. The Washington Post then visualized each homicide on a map, studying arrest rates based on geographical location within each city.

Our objective is to create a web-based machine learning application that can forecast the likelihood of an arrest for a homicide suspect. This prediction will be based on selected factors, including the gender, age, and residential location of the victim, among others.

1. Models

We built the model by processing and transforming dataset from Kaggle.