From arrest to parole, officials commonly use clinical or subjective judgements to make predictions about the likelihood that an individual for whom they are responsible will commit a crime in the future.

In this project, we will use 100,000 synthetic instances from [where to get the data] to conduct a supervised data mining []. Each instance is defined by a unique defendantID, and each defnedantID will has the following fields: gender, race, age, # of criminal convictions, housing, drug convictions(binary), etc. We will predict if the target variable (defendantID) will commit a crime and will be arrested with the following type: Violent felony > non-violent felony > misdemeanor > none. There are 73 features including demographic information like age, race, and category of the crimes in the past, time spent in prison, types and number of convictions, etc. We might not use all of them, but it is good to keep the raw data at first.

Our analysis will provide a useful way to target scarce resources better (provide help to those offenders who are most likely to commit crime again, and hence reducing the number of crimes.) For the Parole department, this model will help them to allocate resources in terms of manpower and budgets to those who will most likely to commit crime again.