**Eileen Patten  
Heinz 95-845: Project Proposal   
March 26, 2018**

**Using Prior Criminal Records to Predict Detailed Recidivism Outcomes**

My analysis will build on my Systems Synthesis project, and comments we received at our midpoint presentation. Our project is being conducted for our client, the Pennsylvania Commission on Sentencing, and will inform the PRS committee of this commission as they consider changes to the Prior Record Score used in PA sentencing to make it more empirically derived.

The current Pennsylvania Prior Record Score (PRS) considers the number and nature of prior criminal convictions and juvenile adjudications. The PRS, along with the Offense Gravity Score (OGS), comprise the sentencing matrix in Pennsylvania, and thus the PRS is one of the two primary determinants of the recommended punishment to be imposed on individuals found guilty of a crime. Offenders are classified into PRS categories either based on their point score (0-5), or through placement into the special PRS eligibility categories RFEL (Repeat Felony 1 and Felony 2 Offender) or REVOC (Repeat Violent Offender).

This current scheme was developed in the late 1970s, with some adjustments over time. While utilitarian notions of public safety, risk, deterrence, and incapacitation appear to have been central to the policy discussion around the PRS, the PRS does not appear to have been empirically derived. Decisions about the PRS scoring system, the creation of the eight PRS categories, and the escalation in punishment for increasing PRS scores seem to have been based mostly on intuition and expertise, rather than empirical analysis aimed to identify and categorize offenders based on the likelihood of reoffending. Under this consideration, the current project will aim to construct one or more PRS alternatives that can serve as an input to improve the sentencing process. The goal of these alternatives is to create categories of offenders that are more consistent with existing data on the likelihood of recidivism and severity of recidivism outcome than current scores.

At our midpoint presentation, we presented findings of a recidivism risk score that measured recidivism within 3 years of being released from incarceration or probation as the outcome variable, as is commonly used in similar tools and analysis. However, our stakeholders expressed an interest in having a more directed outcome – that is, they expressed that they were not so interested in whether any recidivism occurs, which could include minor vandalism, drug possession, etc., but that they were more interested in knowing what type of crime they are likely to commit and, in particular, who was likely to commit serious and/or violent crimes upon release. This analysis is important because it will give us (or eventually, perhaps, judges or other criminal justice actors) information about what type of recidivism is most likely, beyond just whether it is likely at all.

Using the dataset given to us – which includes approximately 131,000 offenders sentenced in Pennsylvania in 2004-2006 who were followed for three years upon community supervision or release from incarceration to determine the occurrence of recidivism (W) – I will attempt to model the type of crime committed in the first conviction post-release (if any). As a first step, I will attempt to model the outcome of whether the next crime is a misdemeanor or felony (Y1). Then, I will attempt to manipulate the data to determine what is the Offense Gravity Score (OGS) of the most serious re-offense within the first 3 years (Y2). The covariates (X) in this exercise are indicators of criminal history, including the types and numbers of crimes committed in the past. Additionally, age minus age of first crime is being used to model “duration of criminality,” as people may become less likely to offend as their criminal career lengthens and they age out of crime. Demographic features such as gender, race, and age will be considered, but may be left out of the model due to constitutional considerations. However, an analysis of fairness of the model will carefully consider the predictions based on these key demographic characteristics.

Because my outcome variable is ordinal (felonies are more serious than misdemeanors are more serious than no reoffence), I will transform the categorical variable into an integer variable for the purpose of evaluation. If I am able to reliably model OGS using the data provided, this will also be an integer score on a scale of 1 to 14. I will attempt several predictive models to see if there are major differences in performance outcomes, but since interpretability of a final model is strongly desired by the clients, I will focus on logistic regression or a single tree as the preferred methods.

Some of the data cleaning and pre-processing will already have been completed due to the stage of the systems process. However, the creation of the OGS variable will likely require a careful manipulation of the data. Additionally, I will attempt to consider the fairness of the model across different demographic groups.

The data provided has limitations in terms of the predictive task. Prior criminal history is available in summary variables, with little to no information about when prior crimes were committed. Ideally, more detailed timestamped information would be available to look at crime-free durations and/or criminal trajectories in terms of frequency or severity of criminal activity. Additionally, the data only provides a snapshot of recidivism through the first 3 years after release, so it is impossible to say who would be likely to recidivate over a longer timeframe. Next, because the data follows subjects who were convicted in 2004-06 for 3 years upon release, many of the most serious criminals who are given longer sentences are not captured in this data. Finally, because of differential patterns in policing and criminal justice, it is impossible to disentangle systematic differences by race, age, and gender that appear because certain groups are more likely to be arrested and convicted than others. For example, despite the fact that marijuana is reportedly used at similar rates by whites and blacks, a black person is nearly 4 times as likely as a white person to be arrested for marijuana possession, according to an ACLU analysis of Uniform Crime Reporting Data and the National Survey on Drug Use and Health.

Please note: because the data has been given to us under a user agreement for the systems project, I will submit the project with synthetic data for testing of my code. My findings will be based on the real data.

**References**

The Pennsylvania Code. Chapter 303. Sentencing Guidelines. Retrieved March 25, 2018, (<https://www.pacode.com/secure/data/204/chapter303/chap303toc.html>).

Solon Barocas and Moritz Hardt. 2017. “Fairness in Machine Learning: NIPS 2017 Tutorial.” Retrieved March 25, 2018, (<http://mrtz.org/nips17/#/>).

American Civil Liberties Union (ACLU). 2013. “The War on Marijuana in Black and White.” Retrieved March 25, 2018, (<https://www.aclu.org/report/report-war-marijuana-black-and-white?redirect=criminal-law-reform/war-marijuana-black-and-white>).