

# CS 221 Project Proposal

Gaeun Kim & Benjamin Anderson

24 October 2019

Define the input-output behavior of the system and the scope of the project. What is your evaluation metric for success? Collect some preliminary data, and give concrete examples of inputs and outputs. Implement a baseline and an oracle and discuss the gap. What are the challenges? Which topics (e.g., search, MDPs, etc.) might be able to address those challenges (at a high-level, since we haven't covered any techniques in detail at this point)? Search the Internet for similar projects and mention the related work. You should basically have all the infrastructure (e.g., building a simulator, cleaning data) completed to do something interesting by now.

## 1 Problem & Objective

- \* Predicting recidivism

- \* Why is this important? It determines people's futures and there's a huge worry with over-incarcerating people because of fears that they might recommit crimes. More information lets us tailor sentencing appropriately to the individual and avoid general judgments that err on the side of "public safety" by imprisoning people who are unlikely to commit crimes again. (In an ideal world with a better criminal justice system not based on incarceration this could even be used to recommend extra help/rehabilitation for people who are at risk of committing crimes again but like we don't live in a utopia, RIP)

- \* "If computers could accurately predict which defendants were likely to commit new crimes, the criminal justice system could be fairer and more selective about who is incarcerated and for how long. The trick, of course, is to make sure the computer gets it right. If it's wrong in one direction, a dangerous criminal could go free. If it's wrong in another direction, it could result in someone unfairly receiving a harsher sentence or waiting longer for parole than is appropriate." (PP)

- \* Problems with proprietary and complex algorithms (see: Rudin)

- \* Objective: Simple, interpretable model that is as accurate as possible (reasonable comparison is current black-box things used in the US, like COMPAS)

## 2 Data & Behavior

- \* Data from COMPAS/ProPublica, also have stuff from New York, presumably if we're getting at some actual things we should be able to use similar methods for both.

- \* Input: set of things about the person

- \* Output: Will they recommit? (or: probability/score to represent on continuum??)

\* distinction between violent and non-violent crime?

### **3 Previous Work**

\* Mention Rudin/ProPublica, COMPAS

### **4 Baseline & Oracle**

\* Baseline: If most people re-commit, predict re-commit. If most people don't, predict not-recommit.

\* Oracle: Could be the actual answer?

### **5 Promising Methods**