

**Machine Learning for Public Policy  
Final Report**

**Syntax Error**

Monday, June 4th, 2018

Submitted by:

Andrew Deng  
Amir Kazi  
Tianchu Shu  
Jessica Song

*With thanks to Rayid Ghani and the efforts of the DSSG*

## Introduction

One in one hundred Americans is behind bars in 2010 - but perhaps more alarmingly the statistic constricts to one in only thirty one if including either on parole or probation.<sup>1</sup> While it may be easy to interpret these numbers are relatively innocuous the truth is that many of these 'ones' are at a significant risk of recidivism. Indeed, though an approximate nine million individuals are released from an American jail each year, 34% of jail inmates were later re-admitted to jail while on probation and 13% while on parole.<sup>2</sup> It is therefore clear that there is a significant issue of recidivism plaguing the jail system of the United States, resulting in a revolving population of inmates who are trapped in a vicious cycle of incarceration, trial and arrest. Among other resultant effects, perhaps the most immediately discernible is the ever growing jail population and the costs associated with interring these individuals.

Johnson County, the most populous county in the State of Kansas with its population of near 550,000 in 2010, is not unique in terms of the issues faced by its jail system and the notion of jail recidivism. Though Johnson County employs a bevy of community correctional agencies, ranging from mentorships, to residential transition, employment training, education, mental health and substance services, familial relations and motivational interviewing, these only reduced the recidivism of released inmates by between 65 and 70% between 2016 and 2017.<sup>3</sup> As such the county was one of the seven out of the seventeen counties in Kansas that failed to meet the state's goal of a total reduction recidivism by 75%.<sup>4</sup>

Of course, this raises the question of why jail recidivism should be reduced, not only specifically in Johnson county but also across jails nationwide. After all, common intuition might question why local communities would want potential inmates reintegrated into their communities. A pressing and self evident answer is the notion of cost. Johnson county is of a comparable size to Bernalillo county in New Mexico, with 575,000 residents to the latter's 675,000, yet each inmate in a jail cell costs Johnson County \$190 to Bernalillo's \$85 in 2015 - making Johnson county as expensive place to jail inmates.<sup>5</sup> More striking perhaps is that the cost of probation is around 4\$, meaning that the cost of jailing an inmate is about forty eight times the alternative.<sup>6</sup> Because the Johnson county jail system is an expenditure of the county government, this means that each taxpayer in the county is paying \$82 dollars in tax per annum in order to subsidize the jailing of inmates.<sup>7</sup> As such, there is a measurable benefit to reducing recidivism and leavening the burden on the county's jail system through increasing the success rate of alternative forms of correction, such as probation or parole.

There is also a societal cost to recidivism, from the inmate's family, to employees of the justice system, to the community that is affected by crime. As such, a reduction in recidivism, by the implicit requirement that it reduces crime, could have significant benefits on the county as a

---

<sup>1</sup> Pew Center on the States, Prison Count 2010: State Population Declines for the First Time in 38 Years (Washington, DC: The Pew Charitable Trusts, April 2010).

<sup>2</sup> Allen J. Beck, "The Importance of Successful Reentry to Jail Population Growth" (Presented at the Urban Institute's Jail Reentry Roundtable, June 27, 2006).

<sup>3</sup> Kansas Department of Corrections, Annual Report, Fiscal Year 2017 (Kansas: The Kansas Department of Corrections, 2017).

<sup>4</sup> Ibid.

<sup>5</sup> The Vera Institute of Justice, The Price of Jails. (The Vera Institute of Justice, May 2015).

<sup>6</sup> Pew Center on the States, One in 31: The Long Reach of American Corrections (Washington, DC: The Pew Charitable Trusts, March 2009).

<sup>7</sup> The Vera Institute of Justice, The Price of Jails. (The Vera Institute of Justice, May 2015).

whole. It is difficult to quantify the exact relationship between a recidivising former inmate and their community, but it is intuitive that successfully re-integrating this former inmate and preventing their recidivism will have positive effects on the county's employment, rate of crime, cost of tax and societal contentment. As such, there are incentives in terms of tangible costs as well more communal effects for Johnson county to tackle the issue of recidivism.

## Related Work

The issue of tackling jail recidivism is a subject of significant discourse in the US. Academics, policy institutions and government agencies have generated a significant amount of literature on the issue, covering a broad range of potential reasons for recidivism and a similarly diverse array of potential policy solutions. Many, such as the National Reentry Resource Center's study on reduction of recidivism, do not focus on specifically identifying inmates for their potential for recidivism.<sup>8</sup> This is particularly noteworthy for our purposes, given that county level government lack the budget to potentially intervene on every prisoner as many of such studies suggest - making their policy recommendations unrealistically expensive at a county level.

Others, like the Data Science for Social Good's own report, do attempt to identify specific inmates to target for intervention but utilize different sets of features.<sup>9</sup> In the aforementioned example of the DSSG's methodology, the project focused largely on health conditions both mental and physical to perform their analysis. Our own feature set is expanded to take into account demographic data, hoping to find links between long theorized instigators of recidivism, such as lack of employment or education, economic duress or familial issues.

## Problem Formulation and Overview

The crux of the problem we are trying to solve is relatively straightforward: Given the resources of the Johnson county jail system, how can we identify which inmates are most at risk of recidivism following their release to intervene upon? This question encapsulates the key aspects of the issue and our solution.

- 1) **Resources.** Johnson county's resources are limited, so our solution must identify a number of individuals to be intervened upon. Ideally, this number should be scalable, such that the county can expand the system should an initial pilot prove successful.
- 2) **Risk.** Given a limited number of interventions, our solution must provide a gradient of predictions as to which inmates are most at risk for recidivism so they can be prioritized.
- 3) **Characteristics.** Our solution must also identify characteristics are observable by the county, ideally prior to their incarceration but at least discernible during their jailing.
- 4) **Recidivism.** In analyzing the existing data, our approach must discern which individuals are repeatedly inducted into the Johnson county jail system.
- 5) **Intervention.** Our solution must provide actionable means to ameliorate the issue of recidivism for the county, either within jail or after release.

---

<sup>8</sup> The National Reentry Resource Center, Reducing Recidivism. (The National Reentry Resource Center, a project of the Council of State Governments Justice Center, June 2014).

<sup>9</sup> Data Science for Social Good, Reducing Incarceration through Prioritized Interventions. (Chicago: DSSG, 2016.)

- 6) **Ethics.** Our solution should keep the dignity of the inmates intact, ideally through providing the means to avoid recidivism rather than with punitive punishment.
- 7) **Privacy.** Our approach should retain the anonymity of the inmates in the data.

## **Data Description**

The data utilized is largely drawn from Johnson County's Criminal Justice and Mental Health departments, described at an individual level.

This is further supplemented by demographic information from the American Community Survey, 2010. This demographic data is not individualized, but instead derived from each inmate's area of residence and unique at the zip-code level.

## **Tools & Methodology**

### **Data Cleaning**

We merged data from Johnson County's Criminal Justice and Mental Health department tables, specifically booking, mental\_health and entries tables using either Dedupe ID or MNI number. We then used repeated occurrences of the MNI number to indicate repeat offenses by the same individual, and used that to create an attribute for measuring recidivism. We also merged this data with the American Community Survey based on zip-codes of each individual to assess demographic factors. We then discretized the data and produced dummy variables where required.<sup>10</sup>

### **Methods**

We first assessed the baseline probability of jail re-entry within one and two years of release. Then, we split the data into training and testing data based on booking\_date. We created 2 separate tuples of training and testing sets based on the Y-label being predicted entry within one year, and predicted entry within two years.

We also used different forms of input for our X variables. We ran our models using 'unbiased' and 'biased' training sets. More specifically, we used:

1. All variables (Using all our variables for training data)
2. Personal & Mental Health related Variables <sup>11</sup>
3. Personal & Bail related Variables
4. Personal related Variables.

We ran these different models to assess the impact of each category of variables, and to see if any variables were 'over-powering' the effect of others.

---

<sup>10</sup> CATS and DUMMIES in jocojims.py

<sup>11</sup> Different Categories and their inputs shown in indpv\_lists.py

For our analysis, we used a variety of machine learning techniques including Random Forests, Boosting, Bagging, Logistic Regression, Decision Trees, Extra Trees, and Gaussian NB.

We used the results from running each model then created a grid to store the results with the best parameters from each model. With each parameter in every model in our grid, we calculated the risk scores for each individual in our testing data and attached the risk scores to a copy of our original data file. Using each individual model, we sorted the data according to the top K individuals with highest risk scores (as predicted by the models).<sup>12</sup> This was done to assess the groupings of the people most likely to reoffend, and to be able to gather additional insights. We conducted this analysis for each of our Y-labels (i.e. re-entry within one year, and re-entry within two years). After carrying out visual exploration to dissect the high risk individuals data for trends regarding particular groups (based on age, race, sex, socioeconomic background etc.), we plotted precision recall curves for each of our individual 'best' models in our grid. After which precision recall curves were plotted for each model, side-by-side for comparison. Lastly, we conducted feature importance to determine the impact of each feature, and to assess the top 20 features impacting recidivism rates across both one-year and two-year return rates.

## **Discussion: Results & Learnings**

### **Using All Variables**

Running our models for re-entry within one year, we saw that while there was little separation amongst the models for precision and accuracy, Boosting was marginally ahead in precision with an auc-roc of 76%. This meant that our model performed better than random. Amongst the top 50 highest risk individuals, we saw that those who were expected to be most likely to re-enter jail, the highest majority had a significantly large bail (of \$10,000), with other local peaks at \$1500 and \$5000. However, this needs to be compared to how frequent certain bail amounts are decided in courts, since it might be possible that amounts of \$10000, 1500 and 5000 are more frequently assigned as bail. Additionally, most of these 50 high risk individuals had received bail type of SUR (compared to PR) and had been unable to bail out of jail (possibly indicating low income and an inability to pay bail). Our analysis also showed that amongst the top 50, a large number of inmates were from Overland Park, or Olathe. This however needs to be weighted, since Overland Park (population: 176,185 in 2011) and Olathe (population: 127,907 in 2011) have significantly higher populations than other comparative areas, such as Shawnee (population: 63,219 in 2011). It was also seen that a disproportionately large number of the top 50 high risk individuals were single (over 80%). Mental health treatment was also a very strong predictor, with over 90% of individuals in top 50 being positive for mental health treatment. It was also interesting to see that amongst these high risk individuals, a large majority were concentrated around the mean per capita income of \$30,000, with fewer individual on either extreme end of the spectrum. Running feature importance, we saw that features most important in our prediction were whether they were referred by someone, if they've had mental health

---

<sup>12</sup> We chose different values of K, such as 50 and 200.

treatment, and if they were bailed out. These indicate a strong influence of mental health related factors as a means of predicting re-entry within one year.

Running our models for re-entry within two years, Boosting remained the best model with auc-roc of 0.75 and had precision of 0.51. We saw that bail amount values had remained the same, with \$10,000 and \$1,500 most common amongst our top 50 high risk group. All individuals in this group has bail type of 'SUR', and were mostly situated in Olathe, Kansas, followed by Overland Park, a reversal from year 1 re-entry predictions. Most individuals had children that were in their grandparents' care for over two years, but less than three. All 50 of the high risk individuals were positive for mental health treatment, a large majority were white (as reflective of the 84% white population in Kansas). Mental Health treatment was the most important feature in our prediction, followed by bailed out.

**Using only 'Personal and Societal' Variables such as marital status, age, children, race, education, rent etc:**

Using re-entry within one year as our Y variable, showed Random Forest as the best model, with an auc of 0.60 and precision of 0.26. Precision at 20% of the population was 0.29. For this feature, we assessed the groupings of the top 200 high risk individuals, and their trends - if any. We're using 200 individuals here to better account for meaningful insights, especially since we have reduced the number of variables in our model. There was a peak count at a bail amount of \$2, 500, with other counts approximately equally distributed between \$1,000 and \$10,000, however not accounting for different weights of the frequency of each bail amount decided by the court. As in the case for models involving all variables, 'SUR' was the most frequent bail type, and a majority of these top 200 high risk individuals were bailed out, and a significant amount of people were single and white. Interestingly however, amongst this group of 200, most people had not sought mental health treatment. Amongst these variables, for feature importance, age, marital status and sex were most important, followed by race. It's possible that race wasn't as prominent due to the significant white majority in this county.

Comparing this with re-entry within two years as our predictor, Random Forest is still better with an auc of 0.60 and precision of 0.33. Bail amount, bail type and bailed out trends were similar to year one re-entry trends. However, there was a significant rise in counts by area, with Overland Park, Olathe and Gardner having the three highest counts by a significant amount. Most people were single in the top 200 with highest risk, and had not sought mental health treatment. A significant difference here was that there was a smaller margin between people who were already repeat offenders and people who had not committed more than one offense in this list of 200. Age, Marital Status and Gender were most prominent features.

**Using only 'Bail and Arrests' Variables such as type of bail, race, arresting agency, gender, case type, type of mental health disorder etc:**

Random Forest was the best model for predicting re-entry within one year, with an auc of 0.71 and a precision of 0.37. For the top 200 at highest risk, most common bail amounts were \$3,500, \$5,000 and \$10,000, with over 90% of the bail types being 'SUR'. Unlike previously, more people were not bailed out. As before, most common areas in the top 200 were Olathe and Overland Park. Most individuals were single, and had not sought mental health treatment. Most

interestingly, most of these individuals were positive for re-entry. Naturally, the most important feature was whether the person was bailed out or not, and what the bail type was. This shows that bail related features are extremely important in predicting re-entry.

For re-entry within two years, Random Forest remained the best model, with an auc of 0.71 and precision of 0.46. Most frequent bail amounts and bail types were the same as above. The most important features remained the same (bailed out, and bail type). There was very little difference in trends between year 1 and year 2 re-entry, indicating that when looking predominantly at bails and arrests, trends remain the same across both years.

**Using only ‘Mental Health’ variables such as type of disorder, type of referral, and other individual factors such as marital status, age and income.**

For re-entry within one year, Random Forest remained the best model with auc of 0.67 and precision of 0.39. Most common bail amounts were between \$1,500 and \$10,000, and bail type ‘SUR’. While there was no change in the top two common areas amongst the top 200 most high risk individuals i.e. Olathe and Overland Park, there was a change in the list of the next few areas, which could indicate that certain factors affect some areas more than others. Importantly, a significantly large number of individuals has sought mental health treatment, and most of these high risk individuals were positive for re-entry. Given the feature selection for this model, the most important feature was predictably regarding mental health treatment. The other important features were the source of the referral and the age.

Random Forest remained the best model for re-entry within two years as well, with an auc of 0.67 and precision of 0.45, the latter of which was higher than that of year 1. Trends regarding common bail amounts, bail types, mental health treatment, race, and re-entry remained the same across the two years. For feature importance, the top features remained the same as well.

## **Policy Recommendations**

Our policy recommendations can be divided into three main phases, targeting the at-risk individuals at different facets of their incarceration.

### *Phase 1: Prior to Entering Jail*

We believe that an ideal approach for intervening on individuals at risk for recidivism begins before they physically enter the Johnson county jail system. This involves identifying characteristics that we believe correlate with recidivism, so such potential risky individuals can be targeted for further assessment once they have entered the jail system as inmates.



Mental health can be difficult to quantify, but given its strength as a characteristic in determining the likelihood of recidivism, it is recommended that the county utilize this feature to identify potential at-risk inmates. Unfortunately, though Johnson county funds a number of mental health resources, it is not guaranteed that every incoming inmate will have utilized these programs or indeed any mental health resources at all. As such, a quantifiable mental health paper trail cannot be relied upon to identify at risk individuals. Other resources, such as police reports on previous arrests or encounters with the individuals, testimonials from family members, or observations during their court appearances may be possible avenues for identification.

An easily observable feature correlated with recidivism is whether the individual is able to pay their bail. The county can identify such individuals with ease and list them for further assessment once they enter the jail system. Most bail in Johnson county observed in our dataset was below \$10,000, a noteworthy amount but not insurmountable with the assistance of bail bonds - so an inability to meet bail may belie the individual's financial difficulties. This system can be further extended to include factors, such as the individual's employment history or lack thereof, area of residence, reliance on social security and even number of dependents as means of assessing an individual's economic vulnerability. Such vulnerability is often indicative of trouble employment history, so given that post-jail employment is a strong measure of reducing recidivism, it is recommended that the county particularly identify such individuals.<sup>13</sup>

### *Phase 2: Within the Jail System*

Once prisoners have entered the Johnson county jail system, the tools available for identifying and influencing individuals increase dramatically.

In the case of mental health, it is recommended that the county utilize its resources to screen as many individuals in its jails for potential mental health disorders as possible, so as to identify which are in possession of disorders with severe behavioural effects for targeted treatment. However, given that studies have identified jails as collection points for the mentally ill in the United States, it is possible that current medical staff in the Johnson county jail is not equipped to effectively treat all the prisoners with risk inducing conditions.<sup>14</sup> Furthermore, ethically, the county should not compromise the treatment of those with less 'risky' conditions in order to devote more resources to those correlated with recidivism. Given that Johnson county jail already has a screening process, putting forward a nurse alongside the booking officer when a new inmate is booked, it may be more worthwhile to devote efforts to improving treatment.<sup>15</sup>

Employment and education interventions are mainstays when it comes to policy recommendations for reducing recidivism.<sup>16</sup> We recommend that Johnson county uses screening and other tools outlined in phase 1 to determine which individuals are most at risk at failing to find employment once they are released. Studies largely agree that employment reduces

---

<sup>13</sup> Tripoli, Stephen J.; Kim, Johnny S.; Bender, Kimberly. "Is employment associated with reduced recidivism?: The complex relationship between employment and crime". (International Journal of Offender Therapy and Comparative Criminology, Issue 54, 2010)

<sup>14</sup> Milgram, Anne; Brenner, Jeffrey; Wiest, Dawn; Bersch, Virginia; Truchill, Aaron. Integrated Health Care and Criminal Justice Data. (Harvard Kennedy School, Program in Criminal Justice Policy and Management, 2018.)

<sup>15</sup> Hendricks, Mike. "Medical Staffing is one of the top problems at the Jackson County jail". (Kansas City Star, 2015).

<sup>16</sup> The Justice Center. Justice Reinvestment in Michigan. (The Justice Center at the Council of State Governments, 2012.)

recidivism, with one study showing that even marginal employment could provide sizeable reductions in recidivism and another stating that former inmates were unlikely to reoffend if able to maintain stable employment during their first year of parole.<sup>1718</sup> As such, given the resources of the county, we recommend that they target those individuals who are at risk of failing to find employment upon leaving jail, especially given the social stigma that limits employment opportunities for those with a criminal record. These individuals should be given treatment that is tailored to their needs and interests, likely ranging from education at a GED level, to employment or vocational training.

Furthermore, we discovered a link between the number of years an inmate was absent from their children's life, measured by the years a grandparents in their zipcode served as caretakers on average. This suggests that familial life may be an important barometer for reducing recidivism. With this in mind, we recommend that the county expend resources to facilitate visitation or other means of ameliorating inmates with their family during their jail time. Indeed, one possibility is to use visitation as an incentive for participating in treatment.

### *Phase 3: Leaving the Jail System.*

Many policy recommendations neglect to examine the impact the post-jail world has on the likelihood of recidivism and perhaps more importantly the importance that community plays in determining whether an individual is likely to return to jail or not. After all, the community is not only a majority stakeholder in the reduction of recidivism but a necessary party for ensuring a former inmate's successful reintegration into the community in question. As such, we recommend the county meet with community organizations, ranging from faith based groups to employers important to the area, in order to help facilitate dialogue as to how to accommodate the needs of former inmates. Ideally, such a meeting would be a relatively budget conscious means of facilitating a bridge for former inmates from jail to civilian life.

Indeed, a radical recommendation would be to shorten the length of post-release probation and parole supervision while simultaneously handing off responsibility to willing community based organizations. A study by the Pew Center suggests that states with a shorter supervision period post-jail have lower rates of recidivism, in part due to a decrease number of re-jailing over technical violations.<sup>19</sup> This has benefits on multiple fronts. It lessens the burden on the state to fund supervision of former inmates. It places the responsibility for rehabilitation on a former inmate on the community's shoulders, the group that ultimately itself must be responsible for their reintegration. Furthermore, it more naturally bridges the gap between institutional supervision in jail to a normalized experience as a member of the community.

In terms of actionable recommendations for the county itself, we recommend that the county expands the impact of its existing mental health services for former inmates. Though

---

<sup>17</sup> Uggen, Christopher. Work As A Turning Point In The Life Course of Criminals: A Duration Model Of Age, Employment, And Recidivism. (American Sociological Review. Issue 65. 2000):

<sup>18</sup> Makarios, M.; B. Steiner and L.F. Travis III. "Examining the Predictors of Recidivism among Men and Women Released from Prison in Ohio". (Criminal Justice and Behavior, Issue 37. 2010).

<sup>19</sup> Pew Center on the States, State of Recidivism: The Revolving Door of America's Prisons (Washington, DC: The Pew Charitable Trusts, April 2011).

Johnson county already funds a number of mental health resources, it is currently not known how well utilized they are by former inmates. It is perhaps advisable that the county more aggressively pursues informing former inmates of the opportunities available to them such that they are more inclined to utilize these resources and hopefully continue treatment for their disorders even after leaving the jail system.

## **Limitations & Caveats**

In our exploration, we weighed the effect that mental health disorders have on an inmate's potential for recidivism. However, mental health is a particularly granular field, among which different disorders range in intensity and behavioural effect. In essence, not every disorder is created equal with regards to its influence on an inmate's potential risk for recidivism and as such it is imprecise to intervene on inmates simply because they have been diagnosed with any mental health disorder. Similarly, the intensity and frequency of the disorder prior to and during the inmate's time in jail likely has an impact on its propensity to affect their potential for recidivism, but would require additional data sources to analyze. Similarly, in many cases it is likely that inmates with conditions that existed prior to their tenure in jail were untreated and undiagnosed, making it difficult to assess how long their disorders have had an influence on their behavior.

Similarly, while we had an intuitive understanding that different crimes, and indeed the gradients of severity present even within a given category, likely had an effect on an inmate's risk of recidivism, we did not specifically seek to analyze which categories of crimes had a correlation with recidivism. This is in part due to the dense legal categorization inherent in the justice system, making it difficult to equate severity of crime without making subjective judgements - at least without significant research into the weighting of crimes by judges, perhaps by length of sentence or some other metric of punishment.

Our analysis of feature importance highlighted that while certain key factors wielded significant weight and staying power over time, less influential features did differ in ranking based on time since release from jail. There was an observable difference in features ranked between 11-20 when comparing one year and two years since release from the Johnson county jail system. It is plausible that feature ranking might change further as more time passes, indeed some studies assess likelihood of recidivism up to five years from an individual's release, but such a timeframe is beyond the current scope of our analysis. Similarly, it is also possible that with a greater breadth of features the rankings would also differ over time.

Our analysis has suggested that most high risk individuals require mental health treatment of some form. Intuitively, it would be ideal for the Johnson county jail system to have the means to identify such individuals before they even set foot in the jail itself. However, people hailing from lower socioeconomic backgrounds are less likely to have accessed mental health resources due to prohibitive costs, social stigma, or lack of awareness. With this in mind, it may be difficult for the county to assess the mental health needs of incoming inmates given a lack of potential data among those who would be most at risk for recidivism.

Another crucial issue faced during our data exploration and processing involved anomalies in the records, largely involving individuals seemingly undergoing rapid changes in their weight or height but also their date of birth and other categorical data. Our intuition suggests that this is likely due to some disparate individuals being listed under the same unique identifier (Dedupe ID), leading to anomalous records and the potential for incorrectly labelling an inmate as having been placed in the Johnson county jail system more than once. As such, we believe that there is a non-negligible number of data entry errors, though we do not believe that the scale is large enough to warrant significant attention.

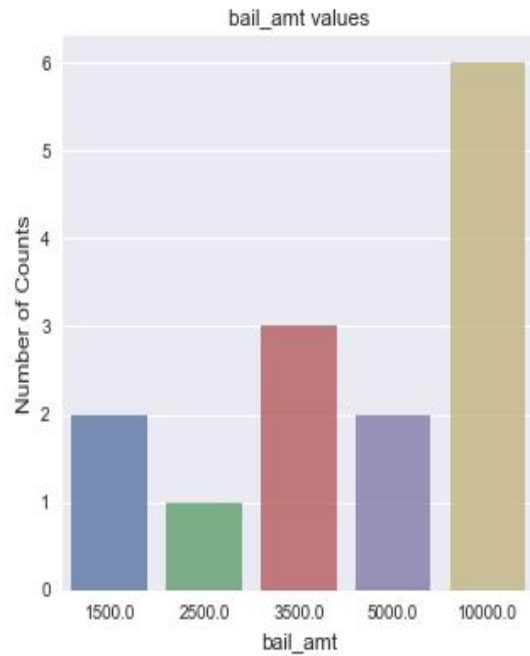
## **Future Work: Room for Improvement**

There is clearly a significant breadth of knowledge to be gained from analyzing which crimes are most highly correlated with potential for recidivism. Potential work in this area could help narrow the scope of interventions and aid Johnson county in specifically targeting individuals before they even set foot in the jail system. Similarly, understanding the nuances inherent in the broad spectrum of mental disorders and determining which disorders are correlated with recidivism offers comparable gains in the potential for more better selection of candidates to be intervened on. Naturally, given the theoretical nature of the work and policy recommendations proposed, it would be remiss to not reflect on the potential that running a pilot program based on our ideas could have on any future analysis.

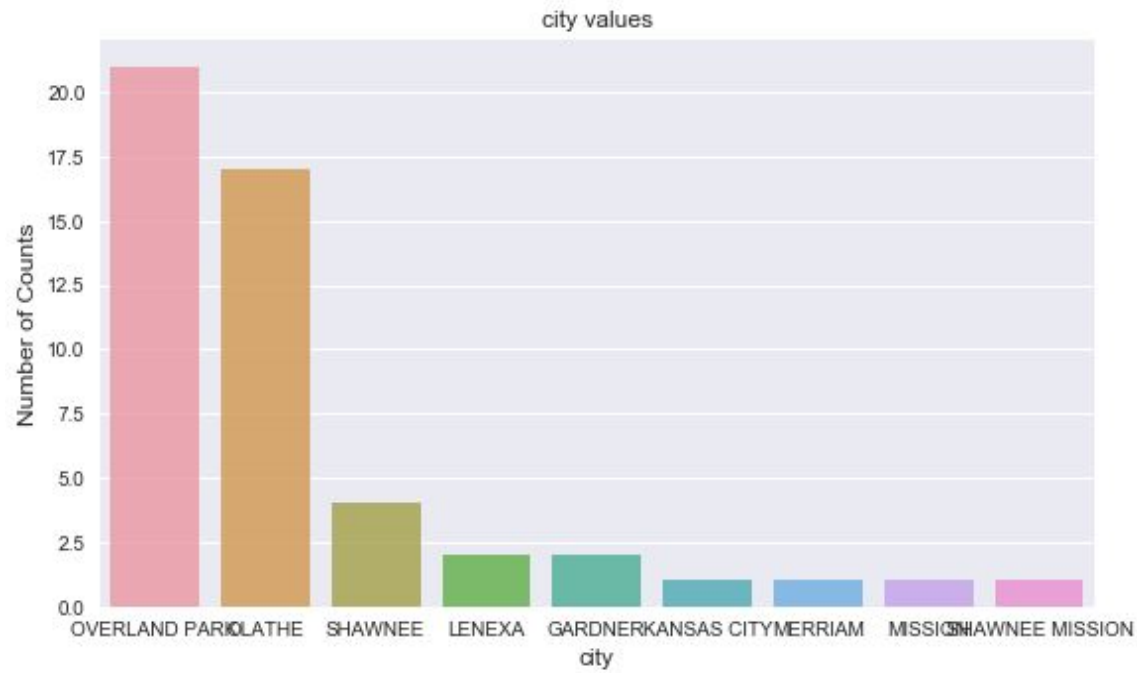
## Appendix: Plots & Features

For brevity of presentation, using all variables in our training data<sup>20</sup>, we saw these results:

### BAIL AMOUNT FREQUENCY FOR TOP 50 HIGHEST RISK INDIVIDUALS, RE-ENTRY IN YEAR 1

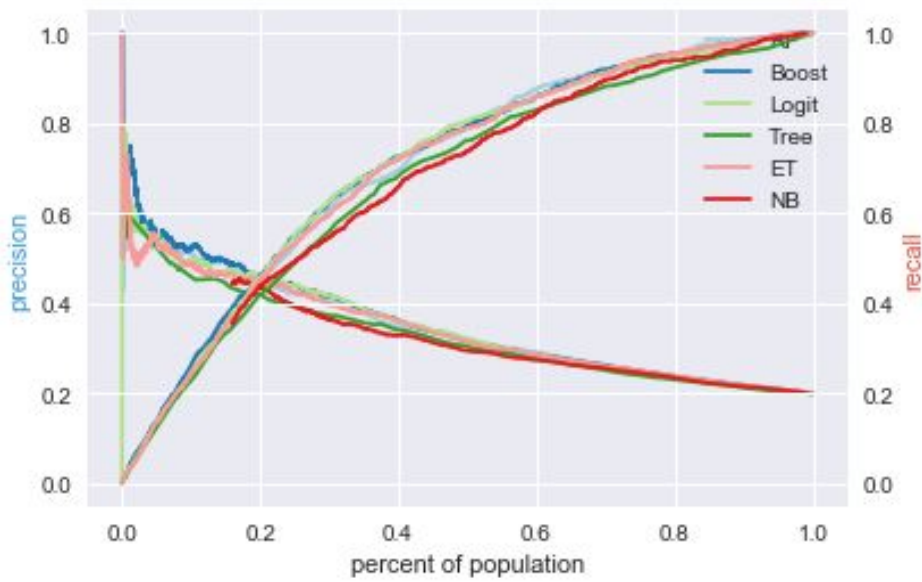


### CITY COUNT FREQUENCY FOR TOP 50 HIGHEST RISK INDIVIDUALS, RE-ENTRY IN YEAR 1

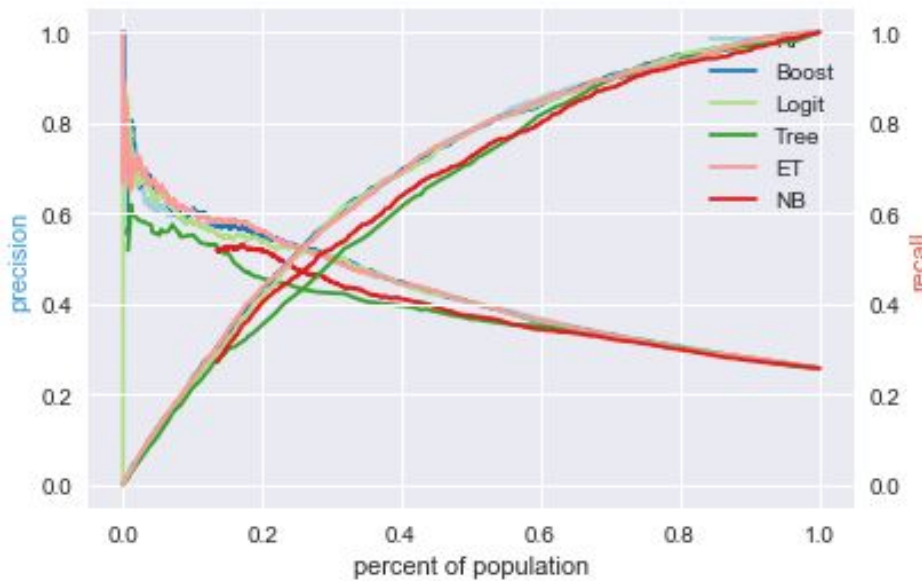


<sup>20</sup> Results for using different combinations (such as personal, or mental health variables) are present in our code file.

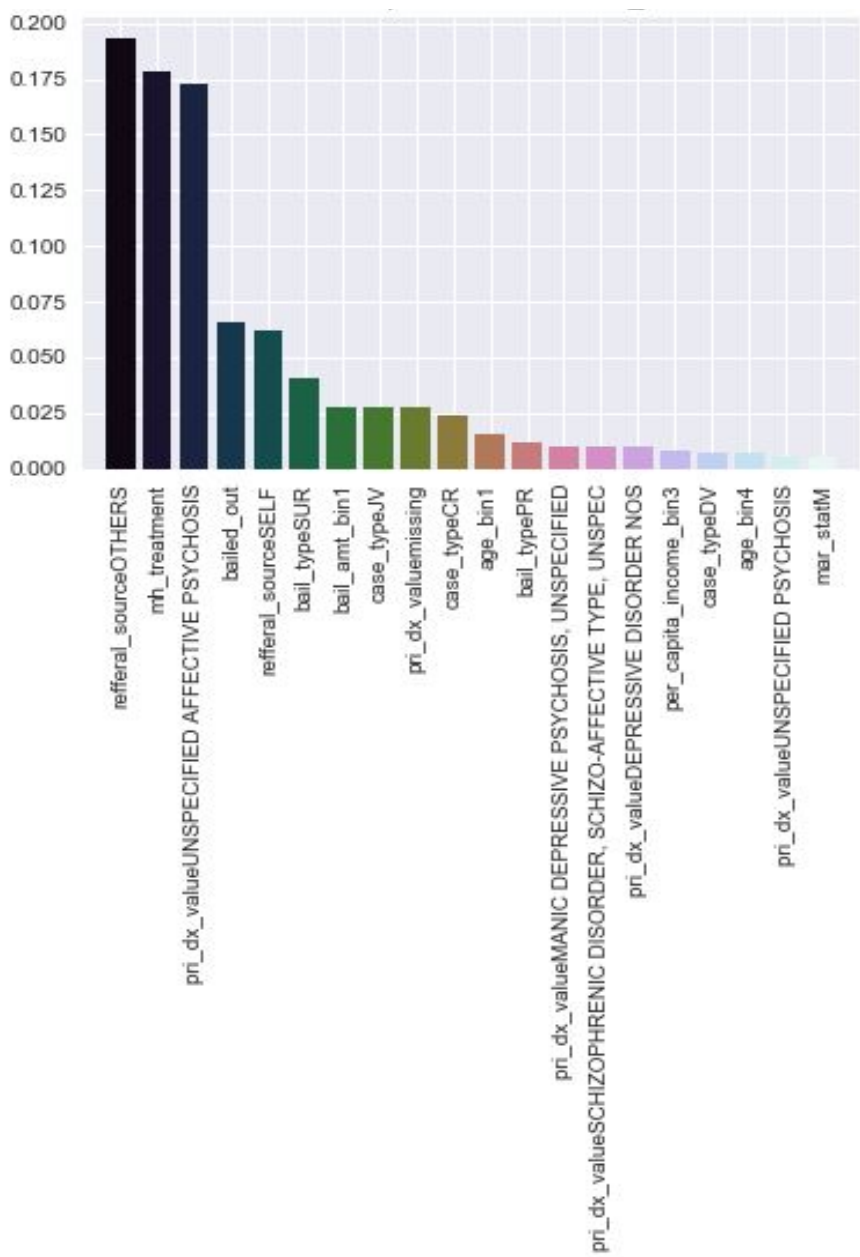
### PRECISION-RECALL CURVE ACROSS DIFFERENT MODELS FOR RE-ENTRY WITHIN ONE YEAR



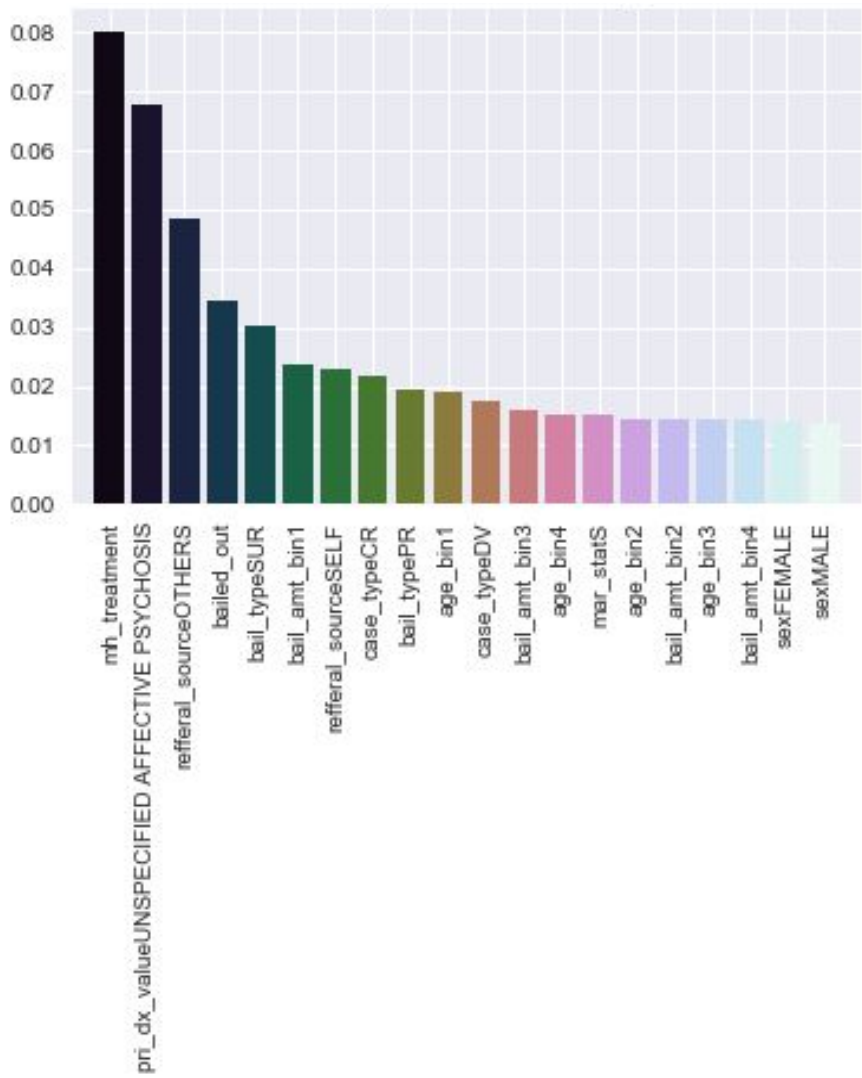
### PRECISION-RECALL CURVE ACROSS DIFFERENT MODELS, RE-ENTRY WITHIN TWO YEARS



FEATURE IMPORTANCE WHEN PREDICTING RE-ENTRY WITHIN ONE YEAR



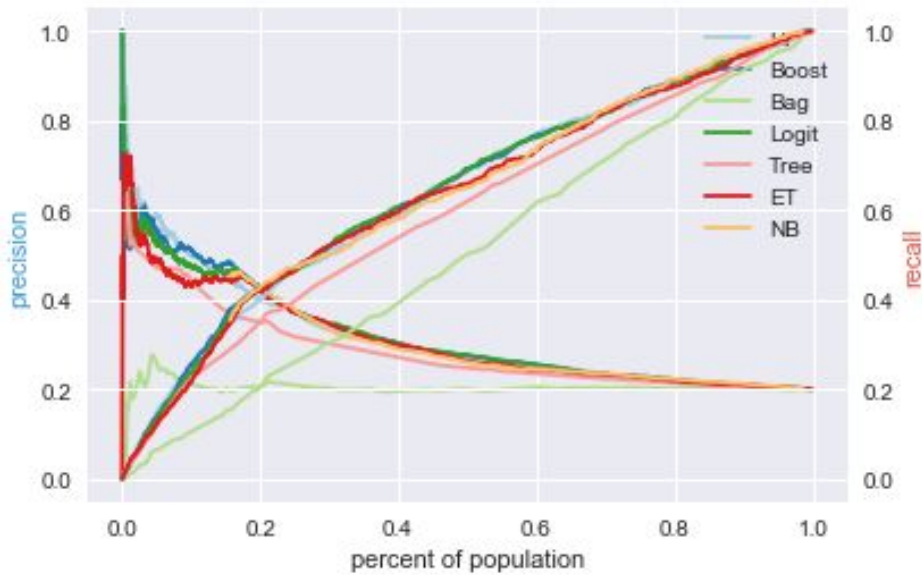
FEATURE IMPORTANCE WHEN PREDICTING RE-ENTRY WITHIN TWO YEARS



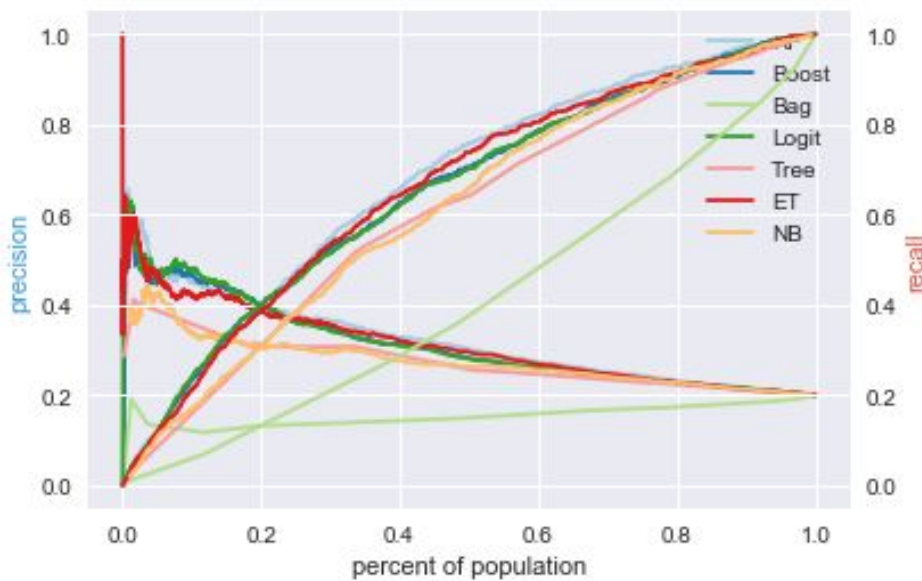


*Precision-Recall Curves for Different Models (Specific Models excluding some variables present in the All Variables Model:*

**MENTAL HEALTH VARIABLES ONLY (FOR RE-ENTRY WITHIN ONE YEAR)**



**BAIL & ARREST VARIABLES ONLY (FOR RE-ENTRY WITHIN ONE YEAR)**



**PERSONAL & SOCIETAL VARIABLES ONLY ( RE-ENTRY WITHIN ONE YEAR)**

