Breaking the Loop

Of Johnson County Recidivism

June 4, 2018

Elena Badillo Goicoechea, Saptarshi Ghose, Loren Hinkson, Natasha Mathur

Background and Introduction

The United States is home to just 4.4% of the world's population yet it incarcerates over 22% of the world's inmates. This system of mass incarceration comes with enormous social and economic costs. According to the US Justice Department, the annual cost to taxpayers of mass incarceration exceeded $81 billion in 2017. This staggering figure is likely a conservative estimate as it only includes the costs of operating prisons, jails, probation, and parole -- while omitting the costs paid by families to support incarcerated relatives. A recent Prison Policy Initiative study estimates these omitted family costs to surpass $100 billion per year -- bringing the total annual expenditure on mass incarceration to over $180 billion per year.

Despite these massive investments, the Bureau of Justice Statistics reports that nearly 50% of all former federal inmates and more than 75% of all former state inmates are rearrested within just 5 years of release. More troublingly, of all incarcerated individuals in local jails in the United States, 64% of persons in struggle with mental illness, 68% have a substance abuse disorder, and 44% suffer from chronic health issues. These metrics point to a disturbing and costly pattern of untreated mental illness -- as well as reincarceration of unwell and vulnerable people. A relatively small number of these highly vulnerable people cycle repeatedly not just through local jails, but also emergency medical services, hospital emergency rooms, shelters, and other public systems -- **with jail becoming the front line for people with complex social and behavioral health issues.** If people with mental health disorders received the services they need earlier, we could potentially prevent situations that would cause them to be arrested (or re-arrested), face new charges in court, and return to jail.

Problem Formulation

Our analysis seeks to identify the individuals most at-risk for recidivism who have had previous encounters with emergency mental health services -- in our case study, within Johnson County, Missouri. As is the case for most local governments, Johnson County has a limited budget with only enough funds to intervene for 200 individuals annually. In the past, Johnson County has relied on expert heuristics and intuition to identify these individuals. We propose a data-driven approach and have developed a series of more sophisticated classification models backed by rigorous analysis of available inmate and mental health data to rank inmates based on their likelihood of returning to jail within twelve months of release given a specific set of feature values. Of those likely to return, we have identified the 200 individuals most likely to benefit from proactive mental health service intervention from the county.

Related Work

Researchers associated with the University of Chicago's Center for Data Science and Public Policy published a 2018 paper, *Reducing Incarcerations through Prioritized Interventions*, that models and analyzes individuals most at risk of being booked in Johnson County jail within the next year, obtaining promising results. Our team met with Erika Salomon, one of the lead researchers for that paper, to gain deeper insight into how their analytical process could inform the way we approached the dataset and built our model. We learned that their analysis also explored mental-health-related features in order to prioritize mental health service interventions for individuals at risk of recidivism in the county. While their analysis makes use of some of the same underlying datasets as our work, their models focus on identifying the 200 individuals that are most likely to be *booked* into jail within the next year as their label. In contrast, our work focuses on identifying those 200 individuals ripe for intervention based on their likelihood of being *charged* within the next year as our label.

Data Description

Data was obtained from work the 'Data Science for Social Good' performed for the Johnson County jail and mental health authorities. It contained information on 80,000 people who had contact with either the mental health system or were booked by the police. The data was split into a series of tables, each of which contained information on demographics, mental health status, diagnoses, details about crimes committed, and other related information.

The above data was augmented with information from the 2010 census. This included data on poverty level, unemployment level, and educational attainment, and was linked to the individuals data by the zip code provided when their booking was processed. Poverty level is defined as the percentage of people living at or below the poverty level in the zip code area, unemployment level as the percentage of the working age population (WAP) that is looking for but without a job, and education level as the percentage of people who have completed a post-secondary degree program.

In the final dataset, each row represents a single charge associated with a booking. The decision was made to use charge level data to better facilitate the creation of all features of interest, including verdict, bail, and crime class detail. Each row contains data about the associated booking, the mental health history of the person who committed the crime, and applicable demographic factors. Ultimately, the data contained about 36,000 charges on just over 20,000 unique individuals. The number of unique inmates was determined using a combination of the 'dedupe_id.' This identifier was developed using fuzzy matching, which meant there were certain entries for which it did not provide a singular match. In order to combat this, we used birthdate as secondary matching criteria. There were other data cleaning issues inherent to datasets involving integration of data sourced from several distinct county level authorities.

Solution Details

After reviewing the dataset and conducting preliminary data analysis, our next step was develop an appropriate data table to use in our model. The first version of the data table we created included columns about every mental health interaction and diagnoses an individual had throughout the entire period for which data was collected, as well as all of their available personal and inmate data. This produced a dataset of over 1 million rows. After further discussion, it was determined that this strategy created repeated rows for the same booking (due in part to variable fields such as weight and height, as well as fuzzy matching used to create unique identifiers), and therefore was not an accurate representation of the subjects we were modeling.

After discussing ways forward with researcher Erika Saloman from the Center for Data Science and Public Policy, we collectively decided to use 'dedupe_id' as a primary identifier for a single, specific individual. Further work around how we joined tables ultimately resulted in a data frame of around 36,000 rows, representing 20,000 individuals -- post-cleaning.

Our feature engineering decisions drove the complexity of the joins required to create the data table. In addition, as the features we used involved data from multiple sources of the database, several of them needed to be type-casted, normalized (in the case of any numerical variable, such as census ratios, age, and number of charges) or dummified (in the case of categorical variables).

We were able to successfully minimize missing values in our final dataset, and then we imputed any remaining missing values using multivariate imputation via chained equations (MICE) via the fancyimpute Python library for compatibility with scikit-learn machine learning classifiers.

The features we used are as follows:

Category	**Type**	**Variable Name**	**Description**
Census Demographics	Normalized Numeric	unemp	Unemployment rate by zip code
	Normalized Numeric	pov	Percentage of the population living at or below the poverty line by zip code
	Normalized Numeric	educ	Percentage of the population who have completed higher education by zip code
Mental Health Diagnosis	Binary	mh_diagnosis	Whether or not they have received a mental health diagnosis
	Binary	anxiety_dummy	Whether or not the person who committed that crime has been diagnosed with an anxiety-related disorder
	Binary	depression_dummy	Whether or not the person who committed that

			crime has been diagnosed with a depressive disorder
	Binary	psychotic_dummy	Whether or not the person who committed that crime has been diagnosed with bipolar disorder or schizophrenia
	Binary	ptsd_dummy	Whether or not the person who committed that crime has been diagnosed with post traumatic stress disorder
	Binary	opps_dummy	Whether or not the person who committed that crime has been diagnosed with oppositional defiant disorder
	Binary	drugs_dummy	Whether or not the person who committed that crime has been diagnosed with a substance addiction
(Most) Recent Mental Health History	Binary	mh_last_365	Whether or not the person in question has had any contact with mental health authorities in the last year.
	Binary	treatment_complete	Whether or not the person completed their designated mental health treatment
Race	Binary	WHITE_dum, BLACK OR AFRICAN AMERICAN_dum, ASIAN_dum	Whether or not the person who committed the crime is of the race in question
Gender	Binary	MALE_dum	Whether or not the person is Male
Marital Status	Binary	S_dum, M_dum, D_dum, W_dum	Whether or not the person is respectively Single, Married, Divorced, or Widowed
Age	Normalized Numeric	booked_age	The person's age on the date when they were booked. This was calculated based on the birth year provided at the time of booking and the booking date.
Criminal History	Normalized Numeric	progressive_charges	The number of time the person has been charged by the date of this booking
	Normalized Numeric	progressive_bookings	The number of times the person has been booked to date
Crime Type Domestic Violence	Binary	CR_dum	Criminal Offense
	Binary	DV_dum	Domestic Violence Offense
Crime severity	Binary	MISDIMEANOR_dum	Whether individual was booked for a misdemeanor

	Binary	FELONY_dum	Whether individual was booked for a felony
	Binary	INFRACTION_dum	Whether individual was booked for an infraction

Given the predefined target of 200 individuals for whom to intervene, we decided to choose between the models created using precision at 1% of the population of analysis (~ 20,000 unique individuals). It is important to note that we chose precision as our target metric because the type of interventions associated with our risk assessment are assistive (as opposed to punitive) and so, incorrectly labelling someone as a "positive" would not be a fatal error. Of course, the 1% threshold could be easily adapted in our model, to produce a flexible "policy menu" for other budget constraints..

Our next decision was regarding the time frames to use when running our selected models. We aimed to select the 200 individuals most likely to return to jail within 1 year, and used a validation period of one year, as Johnson County intervenes on 200 individuals per year. The data provided ran from January 2010 to April 2016. As such, we decided to train models for 1, 2, 3, and 4 years, and to test on the following year for each model instance. This allowed us to select the quantity of information needed to train our model while ensuring the the data was still relevant to our classification problem across each training period.

Over 170 classifiers and specifications were used, including decision trees, k-nearest neighbors, logistic regression, Naive Bayes, along with boosting measures as applicable. The models were then evaluated at a precision of 1% and also against a 'majority' baseline of 12% precision and 50% area-under-the-curve (AUC). Our models were further adjusted to utilize the features and classifiers that had the most predictive power for classifying recidivism. Additionally, after bias testing was completed, additional models were created to see whether it could be further optimized for certain marginalized groups.

Evaluation

In this section we explain our criteria for selecting the most adequate model, out of more than 170 we created, for assessing individuals' risk of recidivism.

As noted previously, the intervention program we are seeking to inform, is assistive in its nature--as opposed to punitive. As such, it is important to note that incorrectly labelling an individual as high-risk, while not ideal in terms of resource allocation, would not be highly detrimental to the individual, which allows us to constrain our threshold as needed to achieve the required precision. Secondly, given Johnson County's budget constraint explained above, we are selecting a very small group for interventions (1% of the analyzed population), so we do not have much flexibility to choose along the precision-recall tradeoff margin. As we could not intervene for all inmates who are likely to reoffend, it it more critical to accurately identify the members of the limited group that could be assisted. Guided by this criteria, we decided to choose the model that maximized **precision at a 1% threshold**.

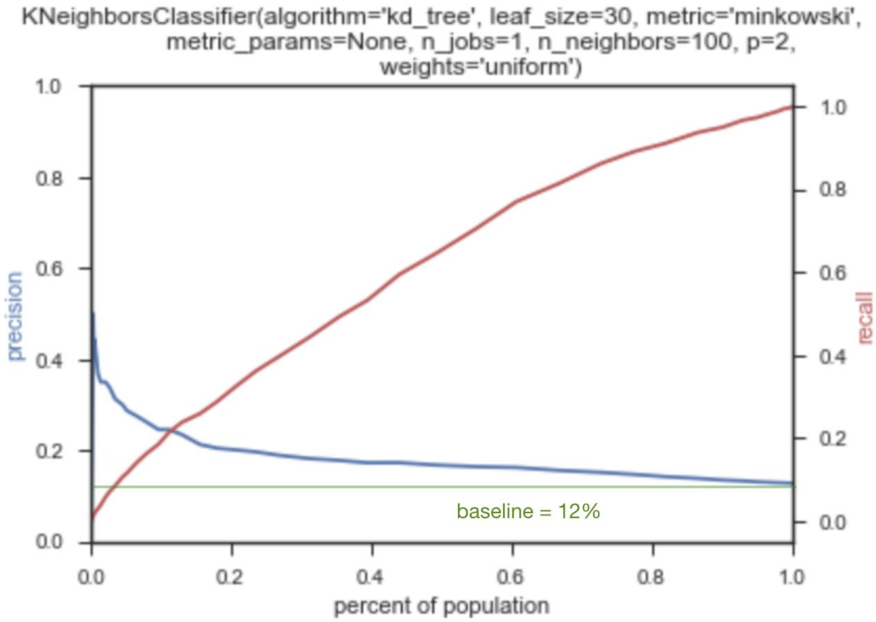
Of course, we evaluated against a baseline of 12% precision and 50% area-under-the-curve (AUC), taking the 'most frequent' class criteria as the baseline, to make sure our model is actually better than costlessly choosing at random. As shown in the results below, our best model indeed was better than baseline, more than tripling its precision rate and exceeding the baseline AUC by 6%.

It would be ideal to also evaluate against experts' heuristics (which would most closely resemble the current approach). Unfortunately, we are not able to simulate their knowledge and instincts with our model. However, as part of the evaluation of our model, we include a **feature importance analysis** that would allow us to contrast and discuss our model with field experts.

Results

At the completion of training and testing phases, all models were assessed as explained above. Keeping in mind a target threshold of 1% , the best performing classifier was a **K-Nearest Neighbors using 100 neighbors**, with a leaf size of 30 and using the Minkowski metric with p=2 (i.e. Euclidean distance). We examined the models that performed similarly well and found that many of them were also K-nearest-neighbors classifiers, reinforcing our decision to use this model. This model achieved a precision at 1% of 38%, more than tripling the baseline.

More detailed results are shown in the graphs below.



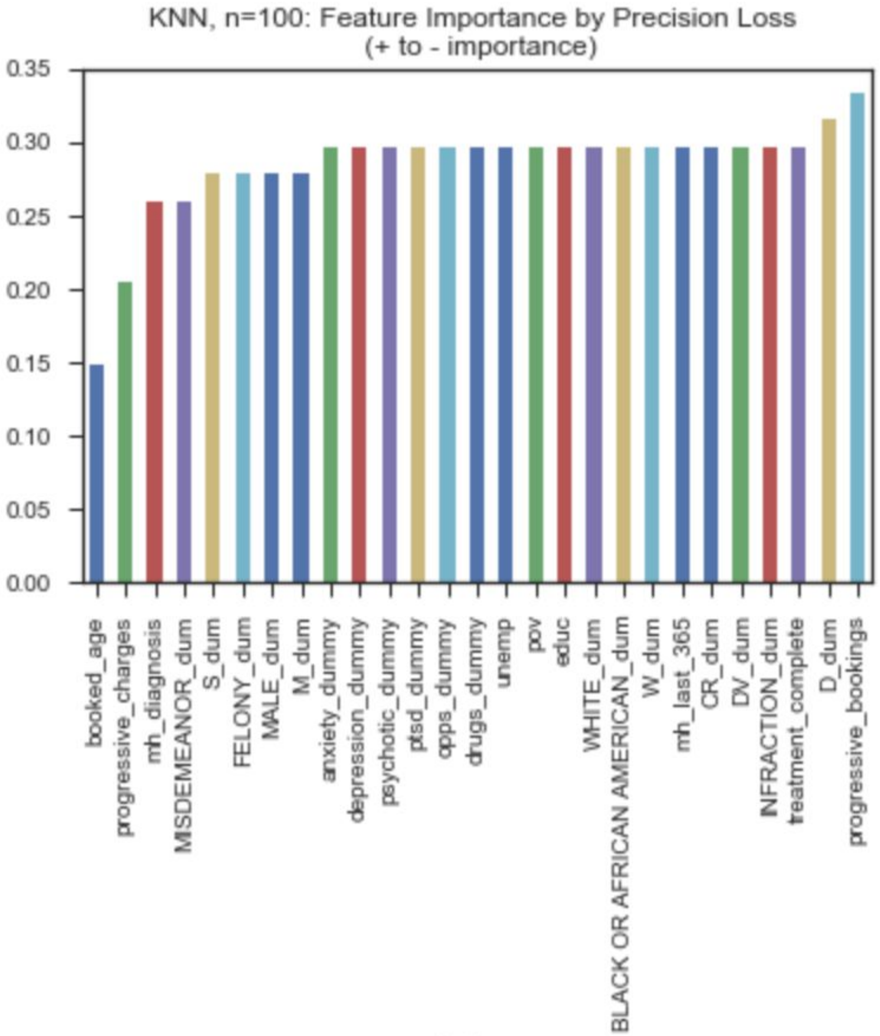
We used this model to select with 200 candidates for mental health evaluation following their release from jail. In the interest of ensuring that we captured all relevant features, we ranked the all the charges by the probability score they received from our model. We then selected the sorted top 200 unique IDs of inmates who had been assigned a mental health diagnosis as the individuals for intervention.

Further analysis was then conducted to determine both the composition of the top 200 list as well as the validity of our model. It was determined that the individuals picked were composed heavily of perons in the 20-25 years age range, and were overwhelmingly male and white. This is not particularly surprising, due to the general demographic makeup of Johnson County. In any case, it is important to acknowledge this observation and discuss possible biases inherent in our data with field experts before our recommendations are deployed in practice. Details about our findings on this point are included in the Appendix.

After estimating the models, we determined that it was critical to assess which of the features we used were most predictive. That information helped us fine-tune our models and detect unnecessary vulnerabilities. Further, because machine learning results should never be interpreted in a vacuum, feature discussion and validation with policy and field experts is crucial for successful deployment. The methodology we used involved computing average precision at 1% via a leave-one-out approach. This process allowed us to see which feature were the most informative for our models. Results of our feature importance analysis for the best-performing model are shown in the graph at the end of this section.

An interesting finding after we analyzed feature importance across most of our models was the salience of both demographic and mental health features. Although the effect of this is muted in the final KNN model selected (in part, we suspect, because of the highly non-linear nature of its classification decision function), feature importances conducted on various models highlighted the demographic factors obtained from the U.S. Census - poverty level, unemployment level, and education level - and the mental health dummies associated with the presence of psychosis and anxiety-related conditions. On this observation, it should be noted that given the lack of more granular data, the census information was obtained using the zip code provided when the individual was booked. This means that general data about the inmates' neighborhood was used -- as opposed to individualized employment data. In any case, it was striking to see how highly determinant an individual's surrounding was to their recidivism outcomes. Some reasons behind this could include a high level of geographical segregation, leading to a lack of viable employment options once the person returns to society. This may be due to the effect that employment opportunities have in reducing the 'cost' to committing another crime -- as well as in stemming the development of adverse emotional/psychological conditions. These factors may play a central role in the vicious cycle we are trying to help individuals avoid.

The mental health findings of our model also proved notable. While many of the factors that affect recidivism are external, as mentioned, the mental health dummies proved to have predictive power as well. Notably, the feature that simply indicated whether or not the person had been diagnosed with a mental health condition proved not to be significant in our analysis. Instead the dummies that showed whether or not an inmate had one of the most common mental illnesses in the United States appeared as an important feature. Based on the DSM categories, the vast number of diagnoses were separated into broad categories. Of these illnesses, the ones that include psychosis or anxiety were the strongest predictors. Further investigation uncovered that while the incidence of schizophrenia or bipolar disorder (the two illnesses entailed in the psychosis dummy variable) is only about 2% in the United States, about 10% of the jailed population in our dataset had been diagnosed with one those illnesses.



Limitations, Caveats, and Future Work

We believe that the models we created showed promising results, especially for the audience we are targeting. They clearly and consistently beat the baseline (by over 3X for our best performing model) and intuitive features proved to be relevant towards our classification problem. However, we acknowledge that before this model (or any other) can be deployed in a real life setting there are certain limitations and concerns to discuss.

As with any data collection procedure, the exact information recorded varied from interaction to interaction. For example, the field for marital status, a standard intake question, included 27 distinct values. When creating the model it became clear that not all of these could be legitimate entries, and we decided to only place weight on those that were marked as single, divorced, married, or widowed. Both the inaccuracies from these atypical entries and the assumptions and judgment calls that we made to address them are sources of additional uncertainty in our models, which could adversely affect their predictive power.

This study takes into consideration mental health data, a notoriously complicated field. There are many factors that go into whether or not someone's behavior is appropriately diagnosed as a mental illness, and even more barriers to to achieving adequate treatment, such as lack of medical insurance, social stigma, and lack of knowledge of common manifestations. Additionally, several mental illnesses, including those identified as important in this model tend to manifest in the early 20s, a group that makes up a large part of our population. This creates a situation in which individuals in the data set who have a mental illness and should affect the model for the overall score may not have a mental health diagnosis on record prior to booking. We began with a goal of determining the inmates who would benefit from mental health treatment upon their release. Over the course of this project, it became clear that predicting which individuals who have not yet been diagnosed will develop a mental illness either during their sentence or afterwards is a task best undertaken with support by professionals with more extensive knowledge of mental health disorders.

Several issues came up around the demographics of Kansas City. About 87% of Johnson County's population is Caucasian, which is not much higher the United States as a whole. Given the limited diversity in this area about 5% of the population in African American. However 20% of the people in our data set were Black or African American. While the disproportionate number of people of color in jail is not an anomaly across the United States, the results from this study may be very different from a student conducted in for example New York City, where less than half the population is white. Therefore the results of this model may not be externally valid.

The data may also not be sufficient to predict for women, as 75% of the jail population is male. This however is not isolated to Johnson County - it is a perennial problem in studies related to incarceration. Due to a variety of social, cultural, and legal factors fewer women are incarcerated than men. This leads to a situation where there is often insufficient data to predict such outcomes for women. The model was run only for women inmates, but performed worse than it did for all inmates as a whole.

In the future, we would like to make use of more of the features within the data provided from Johnson County. There is a wealth of information in the many tables provided that could be parsed more thoroughly. We would also like to see Johnson County codify its results to provide a more accurate picture, and to get information such as education level and unemployment on an individual level, rather than just of the zip code they live in.

Given the limited viability the particular demographics of Kansas City creates, it would also be beneficial to get data from a wider geographic area and more cities. Another important limitation of our model was that in did not account for those who reoffended outside of the Johnson County system, potentially leaving out information that could be used to fine tune the mode and in doing so, labelling someone who did in fact return to jail as a non-returner.

Policy Recommendations

We recommend that Johnson County use our model to generate a list of the top 200 current inmates most likely to reoffend, based on calculated risk scores, and then to provide proactive psychiatric care as appropriate. If Johnson County's capacity to treat grows in the future, the model parameters could also be easily adjusted to provide a larger pool of target individuals.

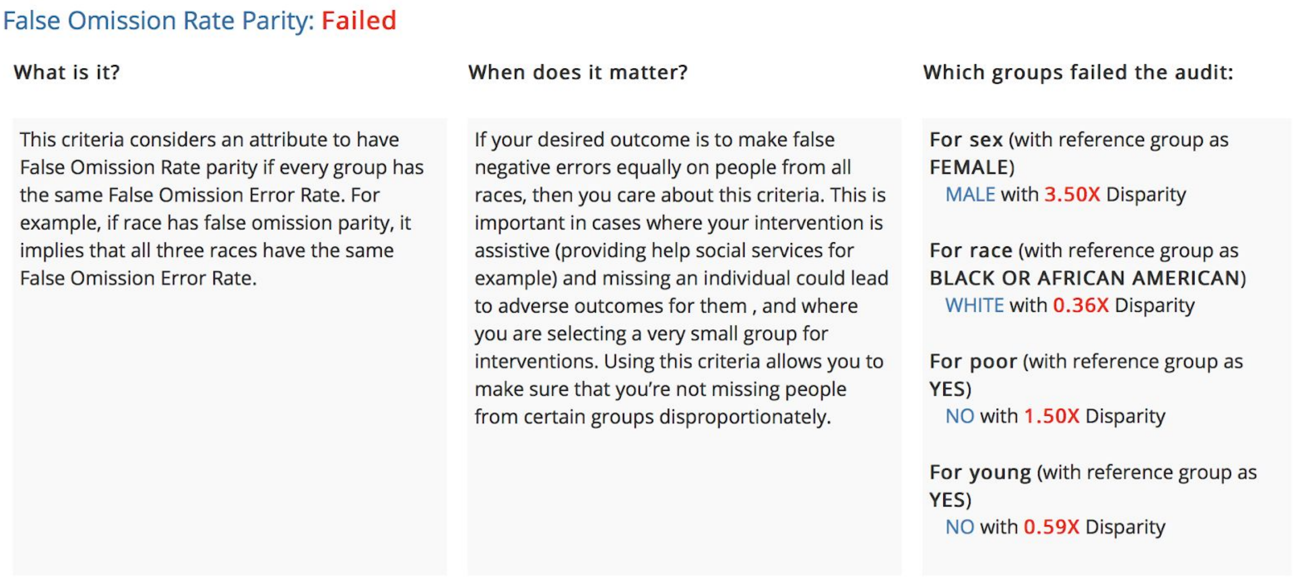
An analysis of feature importance across our 170 tested models showed that education was consistently a crucial factoring in determining recidivism likelihood. As such, we recommend that Johnson County officials also partner with local social service organizations, the public school system, and existing Department of Corrections out-inmate programs to provide targeted educational and vocational training programs. Specifically, we recommend that they make use of the Adult Residential Center Probation Program within the Johnson County Department of Corrections that provides individual-focused intervention programs, including:

- On-site Mental Health Programs
- AA/NA meetings
- Resource Development - Employment Assistance
- Voluntary Religious Services
- Substance Abuse Education and Counseling
- Relapse Prevention
- Pre-Employment Training
- Intensive orientation program for new residents
- G.E.D. and other educational opportunities

As with any model, our results that identify the 200 individuals most at-risk of recidivism in Johnson County should not be used in a vacuum. Instead, Johnson County officials should use our results to inform, rather than dictate, their intervention decision-making. We further suggest that they augment our model recommendations with feedback from decision and policy-making experts in the field.

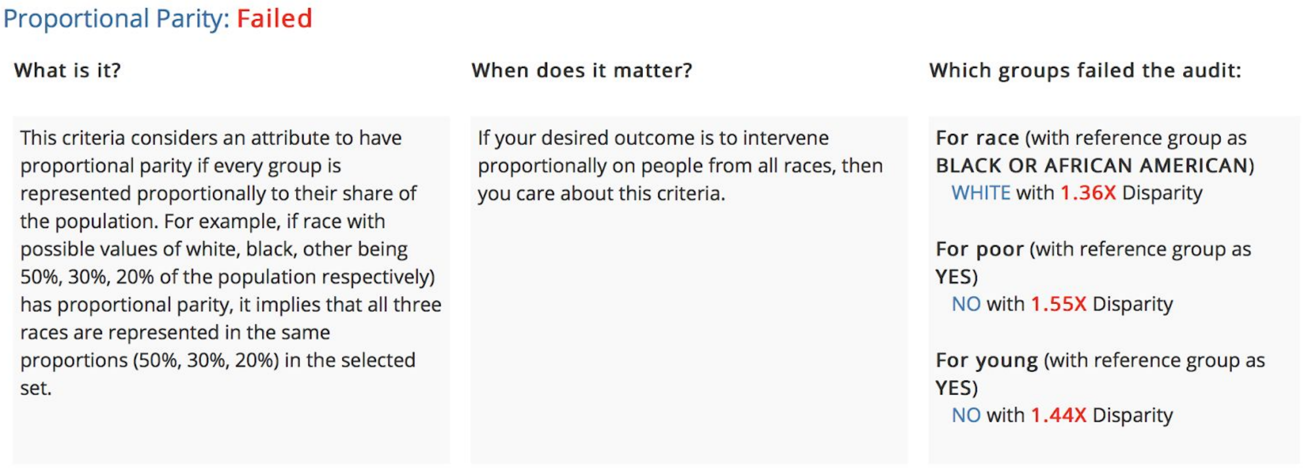
Appendix 1: Bias & Fairness Analysis

Using the bias analysis toolkit developed by DSSG, *Aequitas,* on selected metrics and thresholds, we found that our model's risk scores overrepresented some demographic groups, in the sense that it tended to misclassify as non-risk: 1) poor, defined as living in areas of below average level of poverty , 2) black or African Americans, 3) older individuals (defined as above 25 years old), 4) females. We suspect this may be a result of having less data /relevant feature choice for those groups. More detailed results are shown below:



False Omission Rate Parity: Failed

What is it?	When does it matter?	Which groups failed the audit:
This criteria considers an attribute to have False Omission Rate parity if every group has the same False Omission Error Rate. For example, if race has false omission parity, it implies that all three races have the same False Omission Error Rate.	If your desired outcome is to make false negative errors equally on people from all races, then you care about this criteria. This is important in cases where your intervention is assistive (providing help social services for example) and missing an individual could lead to adverse outcomes for them , and where you are selecting a very small group for interventions. Using this criteria allows you to make sure that you're not missing people from certain groups disproportionately.	**For sex** (with reference group as **FEMALE)** MALE with **3.50X** Disparity **For race** (with reference group as **BLACK OR AFRICAN AMERICAN)** WHITE with **0.36X** Disparity **For poor** (with reference group as **YES)** NO with **1.50X** Disparity **For young** (with reference group as **YES)** NO with **0.59X** Disparity



Proportional Parity: Failed

What is it?	When does it matter?	Which groups failed the audit:
This criteria considers an attribute to have proportional parity if every group is represented proportionally to their share of the population. For example, if race with possible values of white, black, other being 50%, 30%, 20% of the population respectively) has proportional parity, it implies that all three races are represented in the same proportions (50%, 30%, 20%) in the selected set.	If your desired outcome is to intervene proportionally on people from all races, then you care about this criteria.	**For race** (with reference group as **BLACK OR AFRICAN AMERICAN)** WHITE with **1.36X** Disparity **For poor** (with reference group as **YES)** NO with **1.55X** Disparity **For young** (with reference group as **YES)** NO with **1.44X** Disparity