Data Science For Social Good
Summer Fellowship

Evaluating the Effectiveness and Equity of Court Interventions to Reduce Involvement with the Criminal Justice System

Carnegie Mellon University

Nomi Carlebach
University of Oxford

Brian Wesley Hill
Stanford University

Meltem Ozcan
University of Southern California

Liliana Millán
Carnegie Mellon University

Erika Salomon
Carnegie Mellon University

Introduction

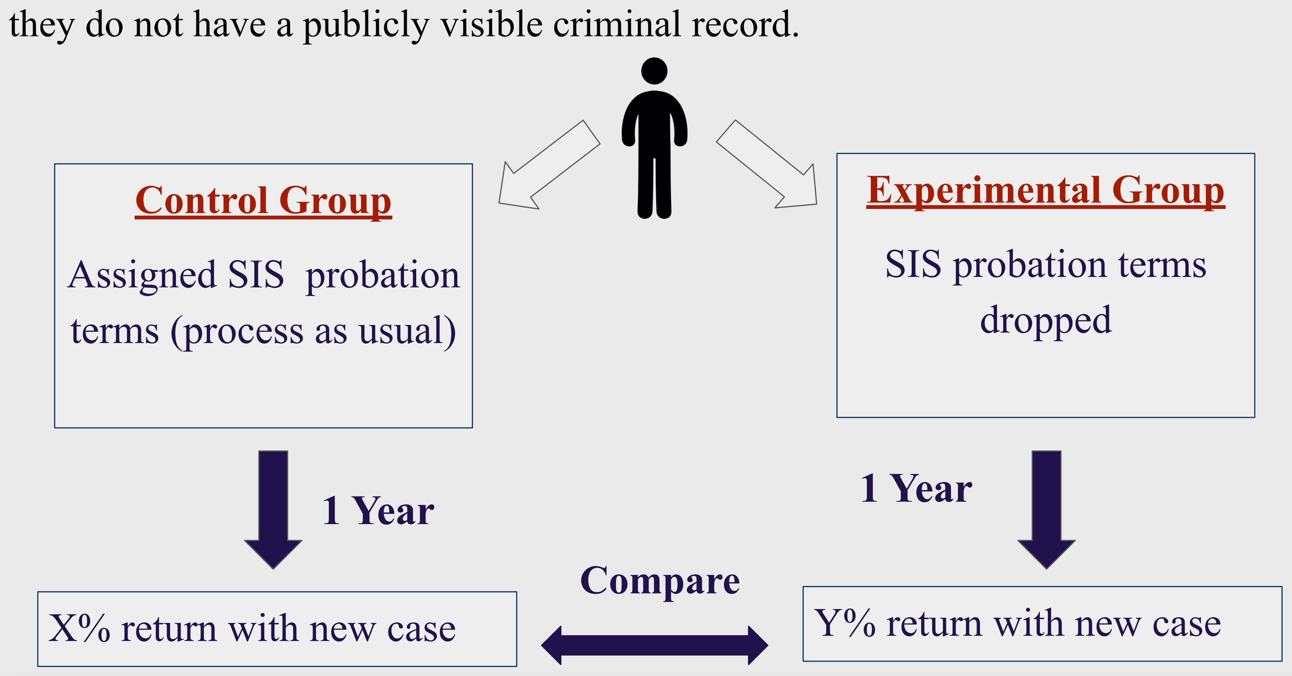
Kansas City Missouri Municipal Court (KCMO-MC) is a judicial circuit court that deals with ordinance violations, such as speeding, trespassing and petty theft, with probation as the most common sentence issued by the court. However, probation terms assigned to probationers, such as community service and anti-theft classes, are often left incomplete and a large number of probationers subsequently return to the court with new cases. In this project we aimed to help the court develop mechanisms to evaluate the outcomes and effectiveness of their interventions in order to reduce individuals' future involvement with the criminal justice system.

Approach

Our approach to tackling this issue was two-fold: 1) setting up an infrastructure that allows the court to experiment with various probation conditions to test the effectiveness of their practices, and 2) building a machine learning pipeline that makes it possible to compare outcomes of pilot programs across different risk groups to evaluate the effectiveness and equity of the program. Here, we focus on predicting the risk of individuals receiving low intensity probation sentences returning to the court with a new case. Together, these components allow the court to determine which interventions work best for which individuals (or alternatively, do not work), and make the necessary adjustments to improve outcomes for the individuals in the system to increase probation completion rates and decrease recidivism. Though we provide a brief overview of the trial design below, the main focus of this poster is the machine learning pipeline which is used to assign individuals to risk categories for comparison of outcomes of the trial across risk groups.

Trial Design

Target Population: defendants with non-violent charges who have been assigned an suspended imposition of sentence (SIS) probation, where if the defendant successfully completes all their probation conditions, their case is dismissed and they do not have a publicly visible criminal record.



Data

KCMO-MC provided us with reports extracted from REJIS and generated through Crystal Reports, which we formatted using a bash script and loaded into a secure PostgreSQL database using SQL. After filtering out cases on violent charges (e.g., child abuse) and excluding cases before the beginning of 2012, our data contained information on 74961 SIS cases associated with 47104 distinct individuals. Multiple probation conditions were associated with each case (e.g., do not obtain similar offenses, 20-hours of community service). Our exploratory analysis showed that all assigned probation terms were completed only for 61% of SIS cases meeting the aforementioned criteria.

Model Pipeline

Features

Demographics
Age, sex, race

Disposition Context
Statute ordinances, charge text groupings (one-hot-encoding)

Temporal data
Days since last (disposition, violation)

Interaction Counts
Number of (cases, dispositions) in the last X years

Target

If within one year of the defendant being put on SIS probation they have an interaction with the police that results in new cases, the defendant is labeled positive. Otherwise, they are considered successful and labeled negative.

Baseline

The baseline model uses only the total number of past cases to generate a risk score.

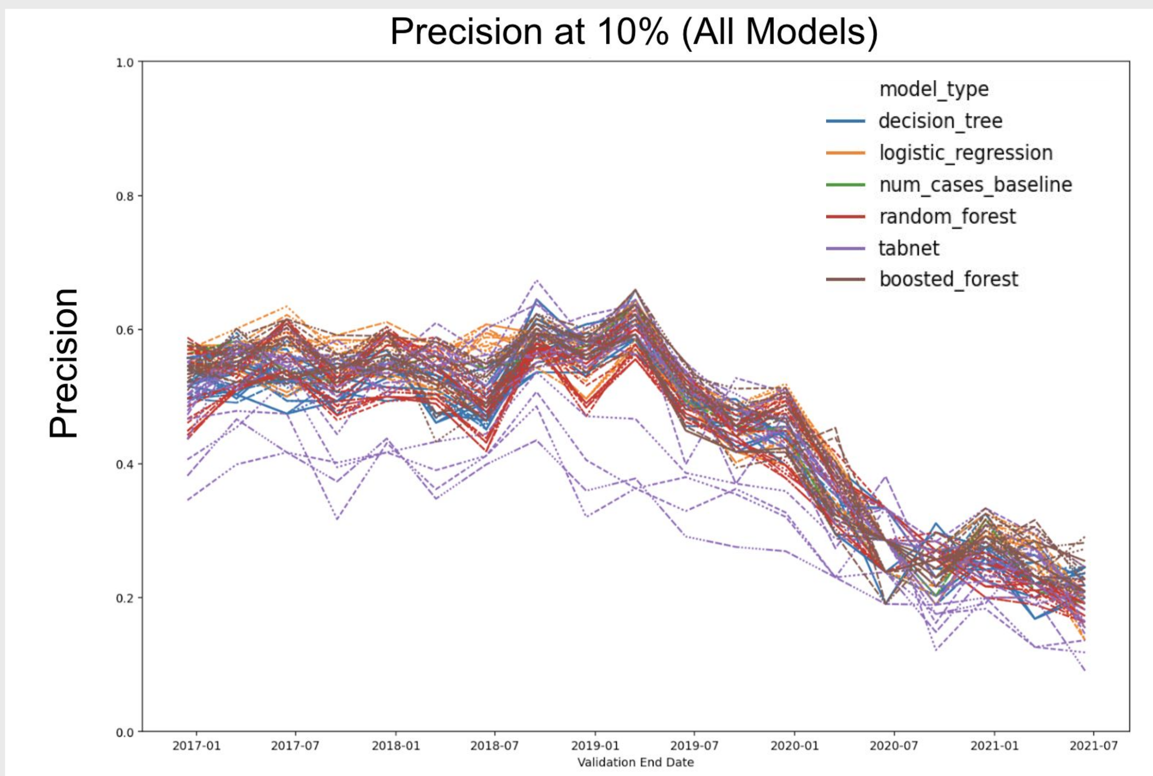
Models

Name	Parameters
Decision Tree	Criterion: Gini, Entropy; Max Depth: null, 2, 5, 10, 50, 100; Min Samples Leaf: 0.01,0.05,0.10
Logistic Regression	C: 0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10
Random Forest	Num Estimators: 500, 1000, 1500; Max Depth: 25, 50, 100, null
TabNet	Learning Rate: 2e-1, 2e-2, 2e-3, 2e-4; n_d: 4,8
Boosted Forest	Num Estimators: 100, 200, 300; Learning Rate: 0.1,0.01, 0.001, 0.0001

Model Validation

We have implemented temporal cross validation instead of randomly splitting training and validation to be as close to the real use case as possible. We have also added wait time after the training and after the validation set to allow for label information to come in, and parametrized the entire process so that the optimal number of years of training data can be tuned as part of the model selection process. In total, we have 10 years of total training data, and have 19 total splits with validation end dates from 2017 through 2021.

Model Performance

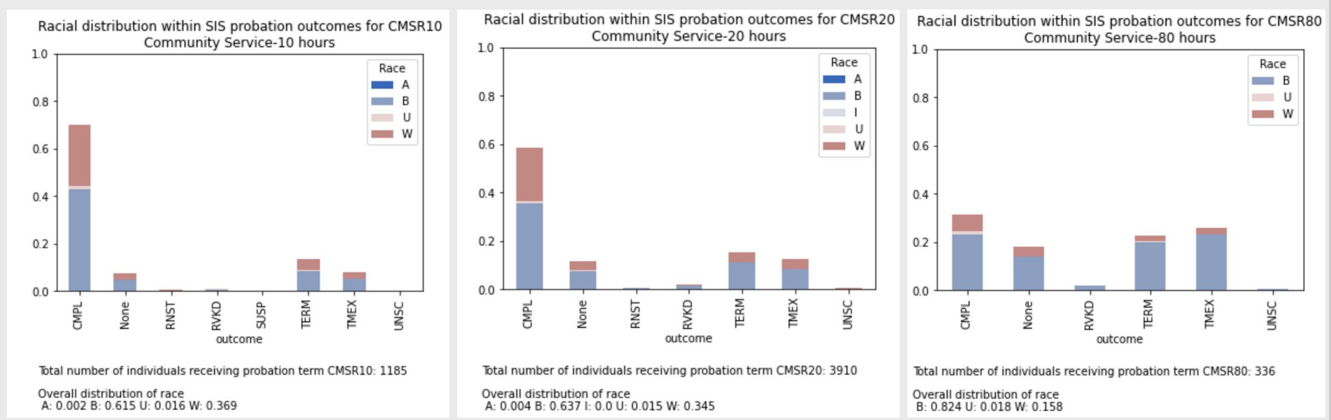


Start of Year	Return Rate
2017	28%
2018	27%
2019	26%
2020	21%
2021	12%

The average percent of defendants returning within one year of disposition date.

Our highest performing models were a collection of boosted forest, logistic regression, and TabNet neural networks. Measured by highest average precision at 10% over all time splits, a sklearn boosted forest with 100 estimators and a learning rate of 0.1 performed best. Models with high precision also tended to have high AUC. The false omission rate which will also be measured over time, and among high performing models we choose the one that minimizes bias. We expect minimal performance tradeoff for reduced bias in outcomes since there are many high performing models.

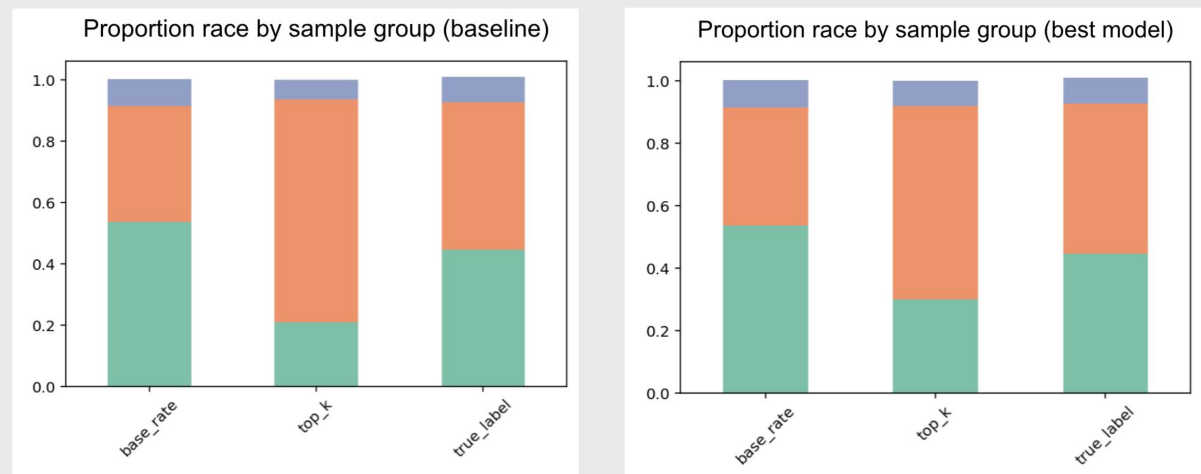
Bias & Fairness



Here we see an example where as community service hours increase, the proportion of black to white defendants increases. Further investigations are required to tease apart the underlying reasons.

	White		Black		Overall	
Top 10%	**Baseline**	**ML**	**Baseline**	**ML**	**Baseline**	**ML**
False Omission Rate	6%	6%	10%	8%	7.3%	6.4%
True Positive Rate	16.7%	26%	27.3%	36.3%	22.3%	33%
False Positive Rate	2.9%	4%	18.5%	14.2%	8.9%	7.9%

Unlike risk models used for sentencing, our risk model will only be used as an analysis tool. Thus, being in the high risk group means that individuals will actually receive more assistance in efforts to reduce their risk of return, and as a measure of bias we are most interested in equity in false omission rate. As expected, the false positive rate is dramatically higher for black individuals than white individuals in both the baseline model and the highest precision at 10% on average over all splits boosted forest model.



We also observe that both the baseline and best ML model put a larger proportion of black individuals in the predicted high risk group (top_k) than actually end up returning (true_label).

Impact

By supplementing the randomized control trial with a machine learning pipeline, we are able to examine heterogeneous treatment effects based on the individuals' risk of accruing new charges and help the court assign probation terms more effectively and equitably, leading to a reduction in recidivism rates. With this infrastructure in place, the court can independently and continuously assess its practices and take the steps to better serve their community. The discussions we had with the judges, prosecutors, and probations staff during our site visit to KCMO-MC revealed much excitement about data-backed solutions, and we have been informed that some judges have already begun taking steps to pivot away from more punitive probation terms after receiving our findings about the court's probation term completion rates. Furthermore, efforts are in place to hire a data analyst to continue the work started this summer.

This work was done during the Data Science for Social Good Fellowship at Carnegie Mellon University.