Project Report

Reducing Recidivism with Targeted Mental Health Outreach in Johnson County

By: Waseem Khan, Colton Lapp, Madi Zhaksylyk

Team MCRT 2

Machine Learning for Public Policy, Heinz College, CMU - Fall 2023

Executive Summary

America's criminal justice system operates at a staggering cost. The financial expense of sheltering and surveilling thousands of individuals is large, but more importantly, there is an enormous cost imposed on the individuals whose lives are uprooted and damaged through incarceration. These damages can have a debilitating effect on people, exacerbating previously untreated problems and leading to a downward spiral of recidivism. As a result, it's imperative to identify and invest in effective preventative measures that can help reduce incarceration rates.

The Johnson County Medical Health Center (JCMHC) seeks to help break this damaging cycle by providing mental health services to previously incarcerated individuals who need help. They have the capacity to perform 100 mental health outreach interventions every month which they hope will reduce the likelihood of reincarceration for individuals by providing critical time-sensitive assistance. The difficulty, however, is in identifying the appropriate 100 individuals to help out of the thousands of previously incarcerated residents in Johnson County.

In this report, we assess the effectiveness of using machine learning methods to help the JCMHC identify which individuals should be prioritized for a mental health intervention. Our ultimate goal is to find individuals who are likely to recidivate in the near future and would stand to benefit from mental health outreach services. Because we don't have appropriate data to measure the causal impact of the JCMHC's intervention efforts, however, we instead formulate our goal as predicting recidivism risk for individuals with mental health needs. More specifically, we predict the risk of returning to jail for two or more weeks within the next six months for all individuals who have demonstrated mental health needs and who were incarcerated within the last 5 years. We focus on predicting incarcerations of 2+ weeks to ensure the JCMHC is helping prevent more debilitating incarcerations, and include demonstrated mental health needs as a necessary condition to try to identify individuals who have a high likelihood of benefiting from the JCMHC's services.

We have access to data from public services linked at the personal level, including data related to bookings, charges, mental health assessments/diagnoses, and encounters with first responders. We use this data to build features that encapsulate a person's demographic identity and unique history with the criminal justice and mental health system at a specific point in time. At any given point in time, there are roughly 10,000 individuals with demonstrated mental health needs and a recent incarceration, and roughly 5% of them will recidivate for 2+ weeks in the next 6 months. We first build a simple ranking heuristic-based baseline model which achieves a precision @ 100 rate of 25%. We then train a grid of ML models and are able to obtain an average precision @ 100 rate of 41% on our best model. We believe our results will generalize to the real world due to our careful validation design strategy which respects the temporal nature of the classification problem. Unsurprisingly, our model shows that features related to an individual's criminal justice history are most predictive. We perform a bias audit and confirm that our model has negligible differences in recall parity for protected groups.

We recommend that the JCMHC incorporate our ML predictions into their outreach program. Before implementation, however, crucial next steps would be to validate the predictive power of the model in the real world and conduct a field test to measure effectiveness by group. The problem formulation should also be revisited and tested to ensure all assumptions are warranted.

Background and Introduction

Johnson County (JoCo), located in Kansas and comprising primarily the suburbs of Kansas City, MO, is the most populous county in the state. The county's Mental Health Center (MHC) offers a range of services including counseling, crisis helplines, inpatient care, criminal justice co-responders, and a mobile crisis response team. These services are essential for addressing the mental health needs of the county's residents.

Untreated mental health conditions in JoCo, as in many communities, often lead to a negative spiral, resulting in repeated periods of incarceration. This has long-term detrimental effects on both the individuals affected and the broader community. A significant proportion of the inmate population in local jails, about 64%, suffers from mental health issues, with 55% meeting criteria for substance abuse or dependence. The prevalence of serious mental illness is notably higher in jails and prisons compared to hospitals, indicating a systemic issue in addressing these needs within the criminal justice system. This situation exacerbates the cycle of mental health issues and incarceration, as the current system is ill-equipped to provide the necessary support.

The JoCo MHC aims to break this cycle by identifying individuals who could benefit from mental health resources and are at risk of future incarceration. Traditionally, the identification of such needs has been reactive, often through screening processes when individuals enter jail. However, the MHC is now developing a program to proactively reach out to these individuals, with case workers aiming to provide care before a crisis leads to re-incarceration. This approach is a shift towards proactive mental health care, focusing on prevention rather than reaction.

Despite the clear need and the innovative approach of the MHC, there are significant limitations in resources. The current budget and staffing levels allow the MHC to only reach out to about 100 individuals each month. This limitation poses a challenge in effectively addressing the widespread issue of mental health and recidivism in the county.

Related Work

One of the current approaches to addressing mental health and recidivism in Johnson County, without using Machine Learning (ML), involves a range of services provided by the Johnson County Mental Health Center (JCMHC). These services include emergency response teams available 24/7, mobile crisis response units, co-responder programs with law enforcement, and specialized services tailored for different groups, such as children and adults with serious mental illness. These traditional methods emphasize immediate support, early intervention, and comprehensive care to address mental health needs in the community.

While the traditional approaches of the Johnson County Mental Health Center (JCMHC) provide critical support, they face several limitations. Firstly, these methods are predominantly reactive, often addressing mental health issues after they escalate. The ability to proactively identify at-risk individuals before a crisis is limited, potentially missing early intervention opportunities. Additionally, these services face resource constraints, limiting the number of individuals that can be reached and treated. This often leads to prioritization decisions that may not always capture those most in need of intervention. Furthermore, without the insights provided by advanced data analysis, interventions might not be as targeted or personalized as they could be with ML-enhanced methods.

Machine Learning (ML) approaches in recidivism prediction represent a significant shift from traditional methods. These models utilize complex algorithms to analyze large datasets, identifying patterns and risk factors that might not be apparent through conventional analysis. By processing data on past criminal behavior, demographic information, and other relevant factors, ML models can predict the likelihood of an individual reoffending. This predictive capability enables a more proactive approach, potentially allowing interventions to be implemented before an individual returns to criminal activity, thereby enhancing the effectiveness and efficiency of recidivism prevention strategies.

The University of Chicago's project in partnership with Johnson County focused on using machine learning to reduce incarceration, particularly for individuals with mental health and substance abuse issues (Matthew J. Bauman et al., 2018). They integrated data from the county’s jail, emergency medical, and mental health services, building a model to predict jail bookings within a year. The model achieved a precision of 51% for 200 individuals, outperforming baselines significantly. In comparison, our project also uses ML but focuses on predicting recidivism risk over the next six months specifically for individuals at risk of longer incarcerations (2+ weeks). We utilize a broader set of data, including a history of arrests and mental health issues, to inform our predictive model. Both projects aim for targeted intervention but differ in their prediction timeframe and specific target group criteria .

Our approach offers several advantages. Firstly, our use of ML enables more accurate and predictive insights, identifying at-risk individuals before they reoffend. This proactive stance contrasts with the reactive nature of traditional methods. Additionally, our ability to process and analyze large datasets allows for more nuanced understanding and targeted interventions. Moreover, the proposed randomized field trial in our project is a novel approach to evaluate intervention effectiveness, setting us apart in terms of innovation and potential impact assessment. This comprehensive, data-driven approach represents a significant advancement in addressing recidivism in Johnson County.

Problem Formulation and Overview

Analytically, we can formulate our problem as follows:

“Every month, for all individuals who have been incarcerated at least once in the last 5 years but are currently not incarcerated AND have demonstrated mental health issues, can we identify the 100 individuals at highest risk of becoming incarcerated again in the next 6 months for TWO OR MORE weeks to prioritize for proactive mental health outreach services?”

The decision to focus on individuals at risk of incarceration for two or more weeks within the next six months was driven by the limited resources available for interventions – specifically, the capacity to conduct only 100 interventions per month. This focus allows us to prioritize cases where our intervention could have a significant and immediate impact. By targeting serious cases, our model is designed to make the most effective use of available resources, addressing the most pressing needs first.

Our formulation is based on the critical assumption that individuals with a history of mental health issues and prior incarcerations are more prone to repeated recidivism. This assumption is rooted in the data available from Johnson County, which includes detailed information on individuals' mental health history and incarceration records. By focusing on this demographic, our model aims to identify those who are not only at a higher risk of reoffending but also those who might benefit most from timely mental health interventions. This approach aligns with the overarching goal of breaking the cycle of re-incarceration linked to mental health issues.

However, this targeted approach has its consequences. By concentrating on individuals with pronounced mental health issues and a history of longer-term incarceration, our model inherently overlooks those who may be at risk but do not fit these specific criteria. This means that individuals who may face shorter periods of incarceration or those whose mental health issues are not documented within the available data might not be identified by our model as candidates for intervention. This limitation highlights the challenges of working within resource constraints and underscores the importance of continually refining our approach to ensure broader and more inclusive coverage in the future.

Data Description

We comprehensively analyzed a wide array of data points from Johnson County to inform our predictive model for recidivism. The datasets included detailed demographic information such as current age, age at booking, sex, race, and marital status. These demographic features were crucial as they often correlate with the likelihood of recidivism and are fundamental for any model attempting to predict such outcomes.

We also delved into socioeconomic data like zip code, residency status, and average income. This information was pivotal as it provided context to each individual's environment, which could influence their support systems and access to resources – both of which play roles in recidivism rates.

Criminal history data were particularly telling, with variables such as the number of previous incarcerations and the duration between them serving as key indicators of an individual's interaction with the criminal justice system. Features like the average time until the next booking and the most common charge type were used to understand patterns of behavior and potential risk factors for reoffending.

We also considered health-related data, including mental health diagnoses, substance abuse flags, and the number of encounters with mental health services. This health data, combined with the frequency and recency of jail bookings, provided a holistic view of an individual's interaction with both the criminal justice and health services systems.

Our data exploration also revealed the importance of judicial decisions and bail history, including the type and amount of bail set in previous bookings, which may affect an individual's ability to avoid pretrial incarceration and could be a stressor that contributes to recidivism.

Risk assessment tools such as the Level of Service Inventory-Revised (LSIR) scores were particularly informative. These scores are designed to evaluate the risk and needs of offenders, guiding intervention strategies. By including LSIR scores in our feature set, we could incorporate a standardized measure of recidivism risk into our model.

Finally, the preliminary exploration of baseline recidivism rates without ML helped us identify the limitations of current approaches and the potential for improvement with our ML model. We analyzed basic strategies such as selecting individuals with the highest number of previous incarcerations or those with the most recent mental health crises and booking history.

By carefully selecting and engineering these features, we could build a model that not only predicts recidivism with higher precision but also aligns with the ethical standards required for such predictive work. Our approach aimed to utilize the predictive power of ML while ensuring fairness and avoiding the perpetuation of existing biases. The exploration of these data fields informed our formulation and modeling choices, setting the foundation for a predictive model that would be both effective and equitable.

Solution

To accomplish our overall goal of assisting JCMHC help people, we needed to establish the ability to both:

1. Identify which individuals are at risk of recidivating
2. Identify which individuals will benefit from JCMHC's mental health interventions, out of all the individuals who are at risk of recidivating

Both of these tasks can be cast as supervised prediction problems. For the first problem, we have over a decade of panel data detailing individuals in Johnson County which we use to identify predictive patterns within. Because we have both people's criminal justice outcomes in our dataset as well as demographic, mental health and other data, we are able to construct features *and* labels for every individual and then employ supervised classification algorithms on the resulting constructed datasets. We take this approach and produce a machine learning model that is able to predict recidivism for our cohort of individuals at a precision rate that far exceeds a naive model.

We are not able to accomplish the second task of predicting intervention effectiveness for people at this moment because we have no ground truth labels in our data that details which interventions are effective for which people. Although the JCMHC may already have data about the people they've helped and whether or not they returned to jail, this data would not be sufficient for us to causally determine intervention effectiveness because we would not have a counterfactual set of individuals to compare results against. It may be possible that JCMHC's current process of selecting individuals to assist may be selecting on features unobserved by us in the data, which would prohibit us from obtaining accurate causal estimates using regression analysis.

For this reason, we recommend that a randomized field trial should be conducted. This field trial would utilize our machine learning model to create cohorts of people with similar levels of risk of recidivating, and then randomize intervention across the cohorts and record the effect size. Doing this would allow us to understand the heterogeneous causal impact of interventions by group. More details about the field trial are given in a subsequent section.

Using Machine Learning to predict recidivism risk: Procedure

We used the open-source ML pipeline package called Triage (originally developed at the University of Chicago's Center For Data Science and Public Policy) to build our training matrices, prototype our ML models, evaluate model performance and audit for bias. Utilizing this system allowed us to iterate through the entire ML process multiple times, adjusting and fine tuning our problem formulation, feature generation process, and bias configuration.

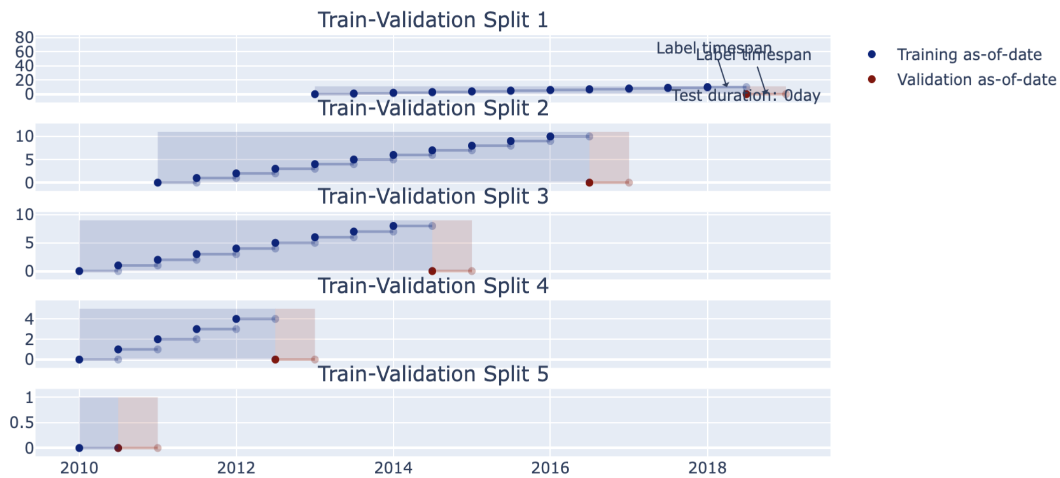
We stored our anonymized data in a secure remote server which our team SSH connected to. We created our ML models by first defining our cohort, label outcome and features in SQL code, and then we fed that code into Triage alongside a .yaml file containing model and temporal specifications. Triage then ran the end to end ML pipeline and saved all the relevant results in our remote database, which we queried for analysis. To streamline our process, we wrote modularized python code to run the Triage pipeline and then fetch the corresponding results, all of which we stored in a Github repository.

The first technical step we completed was translating our relational database into training and testing matrices which we could utilize to build our ML models. We accomplished this by creating a grid of distinct points in time to represent hypothetical model training dates, and then queried data before those points in time to build features and queried data after those points in time to determine label outcomes. We refer to these points in time as “as of dates”, because they represent all the information that was known in the system at that particular time. In a real world deployment setting, the “as of date” would be the current time, and we would be unable to look into the future to discover an individual’s label.

This setup means that for a given individual who was in our data system for a long time, they were represented in our training and test matrices multiple times, but their features and labels were not identical because the circumstances of their life changed between “as of dates”. Thus, each row in our train matrices represented a specific person at a specific point in time. Their history was aggregated into time-invariant columns and their labels were determined by looking into the future and observing if they were arrested for two or more weeks. Our models were judged by their ability to utilize aggregated historical information about a person up until the “as of date” to predict the occurrence of recidivism in the following six months.

Validation Strategy

Our validation strategy was carefully designed with the specific temporal nature of our classification problem in mind. Instead of randomly holding out some rows from training across all of "as of dates" and evaluating model performance on these rows, we ensured that our validation rows were always in the future relative to our training rows. We did this to mimic the nature of the deployment setting and ensure that our past predictive patterns were able to generalize to the future. Additionally, if we allowed our model to train on data from individuals in the future to predict their past, that would be predisposing our model to succeed in a way that wouldn't be possible in a deployment setting.



Because the nature of the criminal justice system and the world more broadly is not static, we also wanted to evaluate how our models' performance would have hypothetically changed over time. To do this, we simulated the evolution of the world over time to our model by iteratively hiding more and more data from the end of our sample. At each iteration of this procedure, we reevaluated our models on the truncated data and recorded the results. This procedure would allow us to hypothetically see if a certain model might perform better before a disruptive event (e.g. a global pandemic) whereas another might perform better afterward. In our final results, we iterated through this procedure 5 times and recorded the model performance in each iteration.

Features

We created informative features for our ML models to use during predictions by aggregating historical information about a person up until an "as of date". The features were created using data from one of three categories: criminal justice information, mental health information, and demographic information. The kinds of aggregations we applied to this information were of the type:

- Count/Sum (ex: count the number of incarcerations a person has had)
- Mean (ex: average number of days for each incarceration a person had)

- Max/Min (ex: maximum value of a mental health diagnosis survey, where 1 = “Positive Diagnosis”)
- Mode (ex: modal value of “race” in bookings database)
- Time since (ex: Time since DOB to assess someone’s age at “as of date”, or time since most recent jail booking)

We also varied the amount of time we considered when we performed these aggregations. For example, we created one feature that counted the amount of incarcerations someone had within the last year as well as another feature that represented the amount of time someone had been incarcerated across their entire life. Varying the time by which we computed aggregations allowed the models to understand the momentum of an individual’s life trajectory and create more nuanced representations of their life. For example, if someone had a high amount of lifetime incarcerations but a low amount of incarcerations in the last year, that would indicate an improving trend, whereas someone else could have a lower amount of lifetime incarcerations but a relatively higher number in the last year. Our nonlinear models would then be able to interact with these features together and use that information for making predictions if it was helpful.

Our features often had missing values that needed to be imputed, either because of data coverage/quality issues or more commonly because not all individuals interacted with the same Johnson County systems. For example, a feature representing whether or not someone tested positive on a mental health screening would obviously be missing in our database for anyone who did not take the screening. To deal with this, we imputed missing values to reflect our intuition of why a certain feature might be missing for someone, utilizing strategies such as:

- Mean imputation
- “Other” string imputation (for categorical variables)
- Zero imputation
- Arbitrarily high value (i.e. 99999) imputation
- In addition to utilizing these above strategies to impute values, we also added a binary “imputation flag” feature that was equal to 1 if a row was imputed.

Model Space Explored

Triage flexibly allowed us to test an expansive model grid for our machine learning prediction problem. We tested different model grids in different stages of our problem formulation, adding and subtracting models to accommodate computational limits and also in reaction to model performance. In addition to varying the models used, we also adjusted crucial hyperparameters to understand their effect on model performance.

Alongside our standard ML models, we also developed two simple baseline models to utilize as benchmarks against our ML performance. When building the baseline models, we strived to create a predictive model that could be maintained with a single SQL pull - i.e. something that relied on a simple threshold or ranking system. The two baseline models we settled on were:

- For all people who tested positive on the BJMHS survey for having a mental health condition, rank them by the number of jail bookings they have in their entire history
- Ranking people by how recently it's been since they were last admitted to jail

The models we trained and evaluated in our final formulation were:

Model:	Hyperparameter varied 1:	Hyperparameter varied 2:	Hyperparameter varied 3:
Random Forest Classifier	n_estimators: [10000]	max_depth: [50, 200]	min_samples_split: [25]
Decision Tree Classifier		max_depth: [1,2,5]	min_samples_split: [25]
AdaBoost Classifier	n_estimators: [300, 1000]	learning_rate: [2, 1, .5, 0.1]	algorithm: ['SAMME', 'SAMME.R']
Gaussian Naive Bayes	var_smoothing: [1e-9, 1e-8, 1e-7]		
MultiLayer Perceptron Classifier	Hidden_layer_sizes: [(350, 200, 50)]	activation: ['tanh']	alpha: [0.01, 0.1]
Scaled Logistic Regression	C: [10, 1.0, 0.1, .01, .001]	solver: ['lbfgs'] … solver: ['saga']	penalty: ['l2'] … penalty: ['l1']

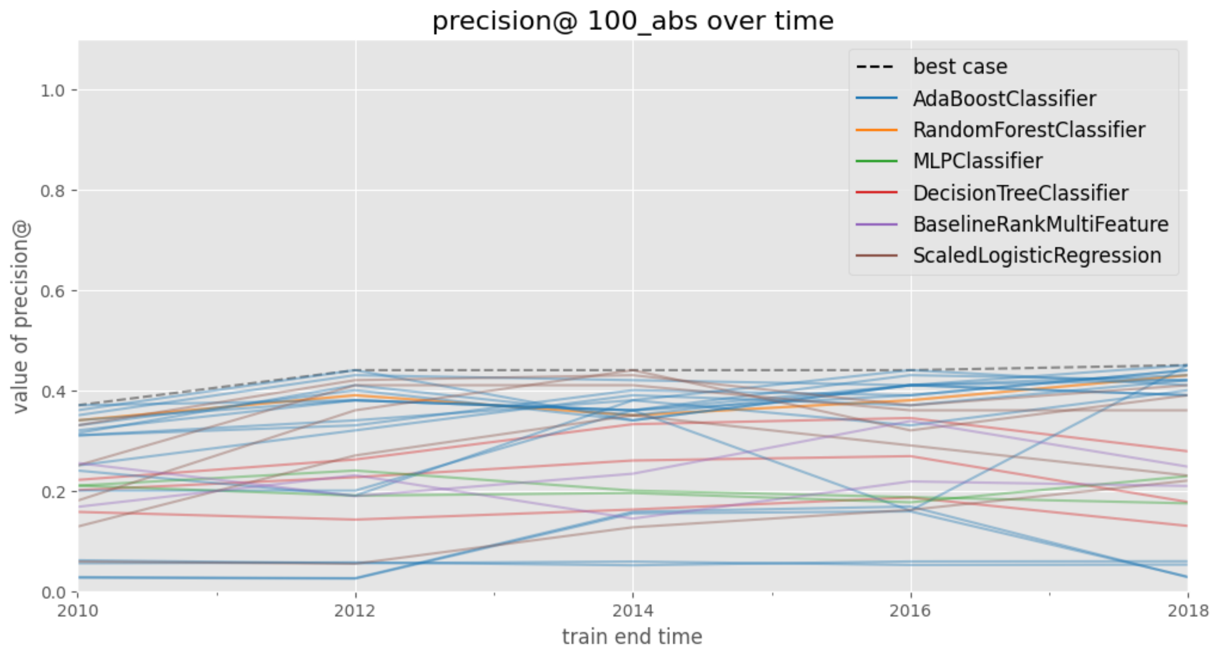
In previous problem formulation, we tried alternative models (such as GBClassifier, SGDClassifier) that we dropped due to poor performance/long train times as well as other hyperparameters (i.e. Random Forest n_estimators = 500) which we dropped because they had worse performance across all specifications and we didn't expect to gain anything from including them. Some of the poorer performing models in the grid were left so that we could compare and contrast their bias performance.

Repository Link:
The code used to create all our results is below:
https://github.com/dssg/mlpolicylab_fall23_mcrt2/tree/main

Evaluation

To evaluate the performance of our ML models, we compared the precision @ 100 across models over time. We believe precision @ 100 is the most appropriate metric because the organization expressed that they only have the capacity to perform 100 interventions a month, and so we think the most important aspect of our models is how many true positives they can return in a list of 100 individuals.

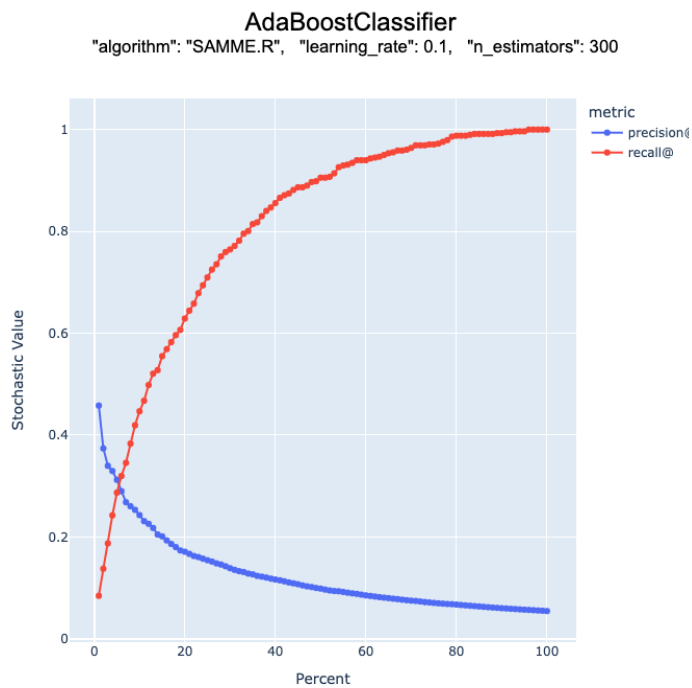
We decided to choose the best model as the one that had the highest average precision @ 100 over our 5 train/test splits. Out of all the models we tried, the ones which performed the best were AdaBoost models, with Random Forest models and Scaled Logistic Regression models coming close for second. Overall, our best AdaBoost model performed 64% better than our best baseline model which ranked by the recency of bookings.

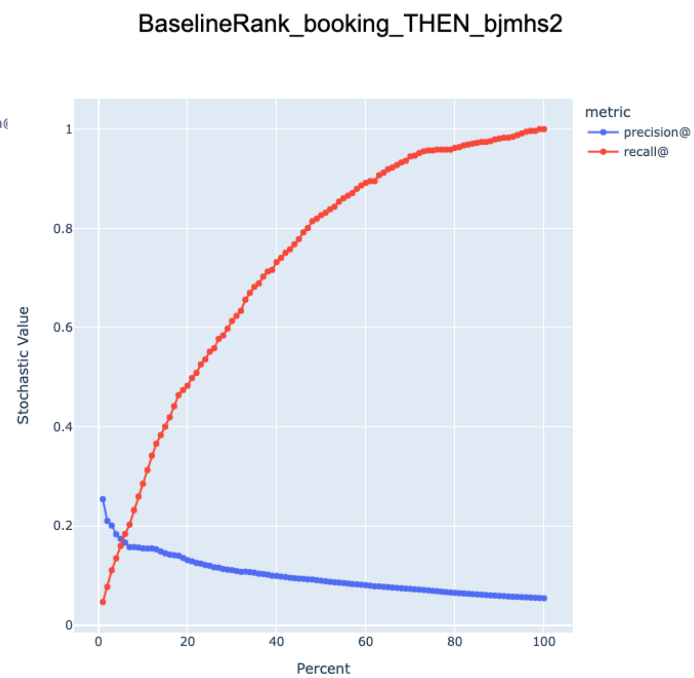


Evaluation Metrics: Best Models and Best Baseline

Model	Model Group ID	HyperParameters	Average Precision @ 100
AdaBoost	20	Algorithm: SAMME.R Num estimators: 300 Learning rate =0.1	**0.412**
Random Forest	2	Max Depth: 200 Num estimators: 1000 Min Samples Split: 25	**0.378**
Scaled Logistic Regression	32	C: 1 Penalty: L1	**0.392**
Baseline: **Rank Bookings**	37	NA	**0.25**
Naive Model: Base Rate		NA	**.05**

Precision/Recall graphs of best model and best baseline on most recent train/test split:





Cross tabs of 10 most important features in **AdaBoost** model 20:

table	feature	Value Top 100	Value **Not** Top 100	Feature importance	Feature Importance Rank
booking	all_**days_since_bookings**_min	65.24	633.757	0.1	1
jail1	all_**avg_incarc_days**_avg	24.487	17.051	0.093	2
lsir	all_**days_since_lsir**_min	447.14	615.171	0.077	3
bkgs	all_prior_count	9.89	4.4	0.05	4
demo1	all_**age_at_bkg**_max	27.02	33.711	0.05	4
demo1	all_**current_age**_max	27.73	35.943	0.04	5
charge	1 year_**dispo_B**_sum	3.13	0.408	0.033	6
jail1	5 years_**avg_incarc_days**_avg	23.848	19.649	0.03	7
charge	6 months_**dispo_B**_sum	1.99	0.182	0.03	7
charge	3 years_**dispo_SS**_sum	0.77	0.19	0.03	7

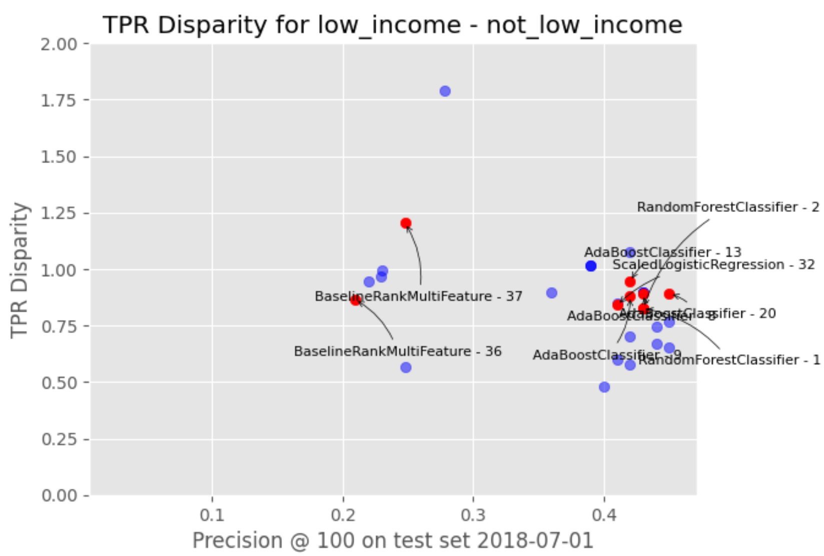
We also conducted a bias audit of all the models we trained. Because our intervention is assistive in nature, we decided that the most important bias metric to analyze is the true positive rate. This is because we want to make sure that our model is able to equally find people to help proportional to their need.

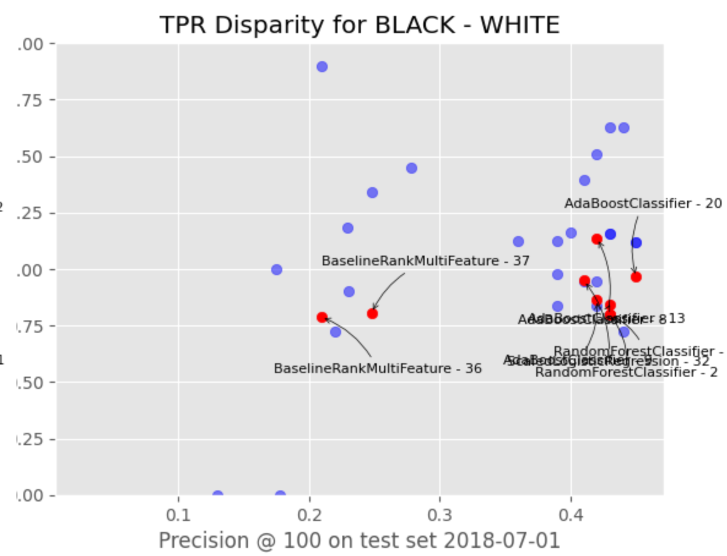
We calculated the bias for three protected groups:

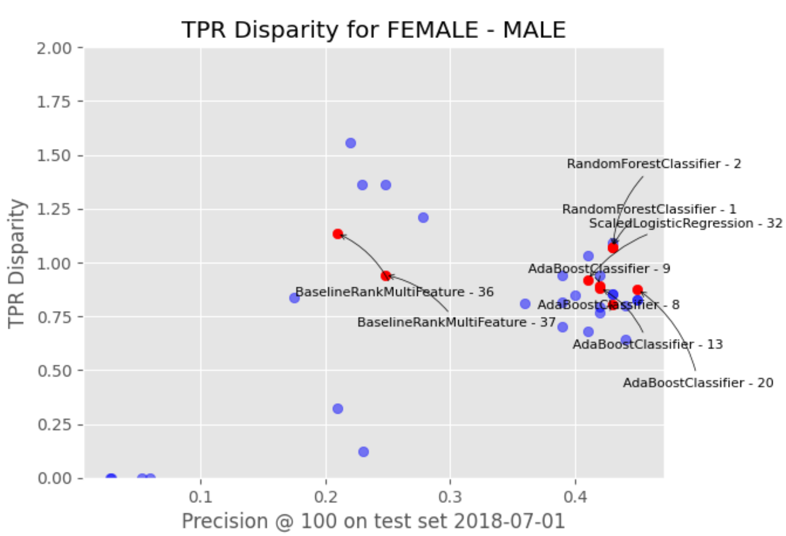
- Black residents
- Low-Income residents (income below Johnson County poverty line)
- Females

Although there are more protected genders and races besides Black people and women, they were not represented in our data enough to give us anything close to an accurate measurement (less than 1%).

Bias Audit Results: all models







Discussion of Results

Our biggest finding is that machine learning is able to provide a significant increase in predictive power compared to a heuristic baseline model. Although ranking individuals by their recency of booking increases precision to almost 5 times the base rate, we are able to get an additional 60% increase in precision @ 100 when we employ preliminary ML models. With further work, we may be able to get this increase in precision to rise even further. Additionally, our results show us that the precision levels of our best ML models are relatively stable over time.

The precision/recall graphs for our best ML model vs our best baseline give us insight into the robustness of the models at higher intervention levels. The precision @ 5%, which is close to precision @ 500 in our case, is 0.31 for our AdaBoost model and 0.17 for our best baseline. This shows that if the Medical Health Center wished to scale up their interventions, the gap in performance would increase somewhat between using ML techniques and heuristic techniques. Still, the baseline precision metrics are not negligible, and could be potentially improved as well by increasing the search space. If the Medical Health Center found it too costly to implement an ML system, they would do well to improve a naive system by implementing a well chosen baseline ranking system.

Our feature importance and crosstabs table reveals that criminal justice history is the most predictive set of features, and that recent involvement with the criminal justice system is especially predictive. This is intuitive, given the fact that we are predicting the probability of returning to jail. It's interesting to note that the model pays attention to the age of which people are involved in the criminal justice system, as well as current age. Additionally, it appears that specific charges are also highly predictive.

We were surprised to see that our bias results showed most of our models being similarly unbiased for all our protected groups. In the future, we would like to investigate these findings and test for robustness by averaging our bias findings across more temporal specifications. One reason, however, that our models may exhibit such low bias is because we are predicting more severe incarcerations (2+ weeks), which may look more similar across protected and unprotected groups.

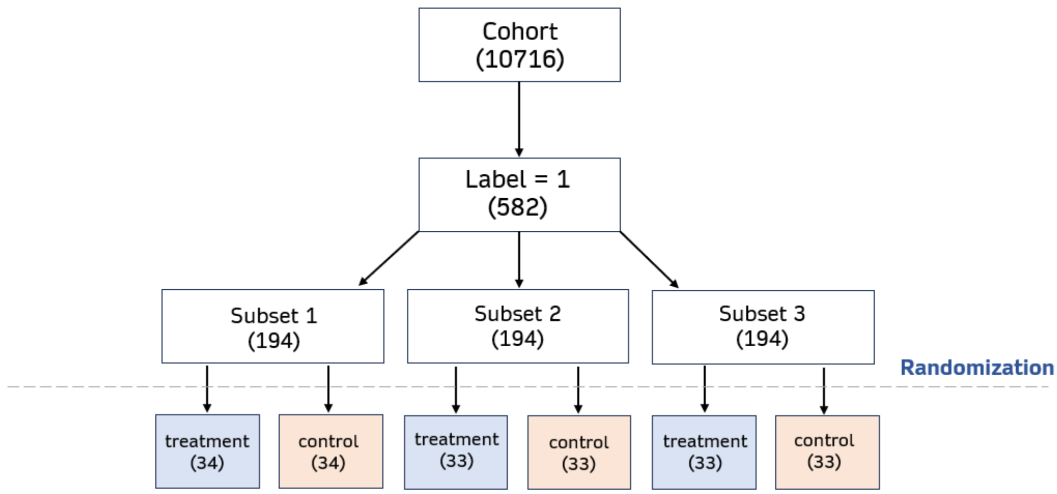
Field Trial Design

To validate the accuracy of our model on truly unseen future data, a field trial could be implemented over a 1-year period. It would provide a dynamic evaluation of the model's performance, offering the opportunity to observe trends or shifts in accuracy over time (12 months). Also it would ensure a robust evaluation across different time periods to confirm the reliability and consistency of its predictive power.

Also, we want to assess the effectiveness of the intervention on targeted individuals. This can be achieved by randomizing the treatment assignment to the high-risk individuals with (predicted outcome = 1) using a tiered approach. In our case, individuals predicted to be re-incarcerated for more than two weeks in the next 6 months can be divided into subsets based on their scores. From each subset an equal number of individuals get randomly assigned the treatment and control groups, while ensuring the total number of treated individuals is 100.

Eventually, outcomes for treated and untreated (control) individuals can be compared across every subset. Assuming randomization was carried out accordingly (no selection bias), the differences in the outcome (re-incarceration) between the treated and untreated individuals should represent some measure of intervention's effectiveness. Concepts like sample size and statistical power should be meticulously considered to ensure robustness and validity for causal inference.

Presented below is one possible design of a randomized control trial to evaluate an intervention for our cohort (numbers for the best model on the most recent train-test split).



Additionally, collecting qualitative data through interviews with treated individuals might provide insights into the personal impact of the intervention, such as improved mental health, increased access to services, and any perceived reduction in behaviors that could lead to re-incarceration. This qualitative analysis could provide insights into the specific aspects of the mental health services that are effective or require improvement, allowing for adjustments to enhance the intervention's impact. The success of the outreach program could be further analyzed by stratifying the results by demographics to ensure that the intervention is effective across all groups and does not inadvertently increase disparities. This stratified analysis would also provide an opportunity to adjust and tailor the outreach services to different subgroup needs, thereby improving overall program effectiveness.

Policy Recommendations

Based on our comprehensive analysis, model development, bias audit, and the proposed field trial design, we offer the following policy recommendations for Johnson County's Mental Health Center:

1. Validate the model's accuracy in real-world conditions by comparing the predicted outcomes against actual re-incarcerations. Also, ensure the model is similarly accurate in predicting outcomes across different subgroups of population (based on race, sex, income). Adjust and/or refine the model as necessary based on the findings to maintain or improve its predictive accuracy.

2. Conduct a randomized controlled trial to assess the effectiveness of targeted mental health interventions on reducing re-incarceration rates. Use a tiered approach to randomization, ensuring that individuals across different risk levels are included in both the treatment and control groups. Evaluate the outcomes to measure the direct impact of mental health services on recidivism across diverse groups, categorized by gender, race, and income status. The purpose of these evaluations is to ensure that the intervention is producing the desired outcomes and that it does so equitably across all segments of the population. Collect qualitative data by facilitating feedback from the participants on their personal experiences to inform potential improvements in intervention.
    i. In the event that the intervention is found to be ineffective, revisit the design and strategy of the intervention. This may involve redefining target groups, recalibrating service delivery methods, or introducing new types of support to better serve the needs of the individuals.
    ii. Should the evaluation reveal disparities in effectiveness—indicating that the intervention does not benefit all groups equally—additional predictive models should be developed based on effectiveness of the treatment. These models would aim to understand and predict which factors contribute to the intervention's success or failure across different demographics.
    iii. If the field evaluations indicate that the intervention is equally effective across all groups, then the current model can be maintained and fully implemented. This scenario would suggest that the intervention is functioning as intended and is contributing to the goal of reducing recidivism equitably.
3. Establish a feedback loop where data from the real world is used to inform model updates and intervention strategies. Encourage a culture of continuous learning where insights from data are leveraged to enhance service delivery and policy decisions.
4. Explore potential partnerships with local organizations and stakeholders to expand the reach and impact of mental health services.

Data Limitations

While the model offers promising insights, we acknowledge that our analysis was constrained by data limitations, such as potential inaccuracies and the risk of biased measurements from untrained individuals. For example, 0.4% of individuals had 2 or more different values recorded as their race at different times in the jail inmates history data.

Data coverage for some important features was incomplete: Brief Jail Mental Health Screen results, which we believed to be one of the key predictors of the outcome, are available only for inmates booked after November 2016; income status was accessible only for previous inmates who have interacted with the Mental Health Center.

The model's predictive capabilities are based on records in the Johnson County services databases, that is, we had information only on individuals who previously interacted with the county's public entities. This implies that we may be overlooking previous inmates whose mental health status is undocumented due to several reasons, including failure/reluctance to seek assistance and/or recent in-migration.

Caveats

Our predictive model, while robust in the testing environment, requires validation in real-world conditions to confirm its predictive accuracy. This step is crucial as it serves as a safeguard against potential shortcomings in model construction that could be exposed when applied to live data. Furthermore, we need to ascertain the impact of mental health interventions on reducing re-incarceration across various groups of population. A carefully structured field trial should be employed to examine the effectiveness of these interventions, ensuring that our approach is not only efficient but also equitable and effective in achieving the desired outcomes.

Future Work

To enhance the model's robustness and applicability, we would work on several adjustments and improvements to our model. First, in collaboration with stakeholders we would review and refine the model assumptions, including the cohort definition (Why booked in the last 5 years? Why with previous signs of mental health challenges? Why re-incarceration for 2 or more weeks?) and predictive window (Why 6 months?).

Next, we believe we could improve the predictive power of the model with a deeper understanding of the existing data and, consequently, expanding the feature set to include specific details of cases, charges, arrests, and bail history.

Another way to develop the feature set is by incorporating new data from other services, including details about education, employment, use of public services, and census information, to provide a more holistic view of an individual’s life.

Appendix

Triage Configuration File:

You can access the Triage configuration file at: https://github.com/dssg/mlpolicylab_fall23_mcrt2/blob/main/Triage/config_files/run_yaml_files/waseemk_11_52PM__05_12_2023.yaml. It includes the labels, features, and model grid used in our latest model run. It also includes the bias audit configuration at the end of the triage configuration file. Each section is well documented with relevant comments.

The definition of our labels are fetched from the SQL file located at: https://github.com/dssg/mlpolicylab_fall23_mcrt2/blob/main/Triage/sql/labels/assignment_3_labels.sql. The SQL file contains detailed comments on how we decide our labels. The same is reproduced here for ease of reference:

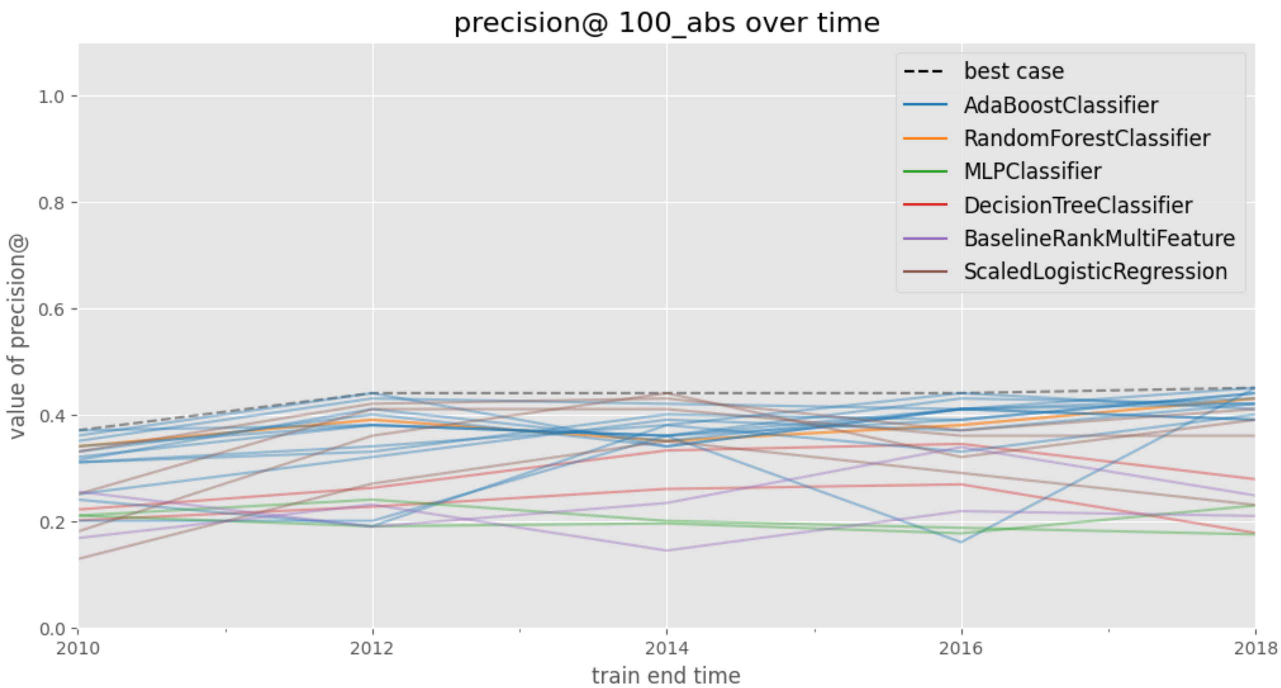
"This SQL query is designed to identify individuals at risk of recidivism within Johnson County. The query selects inmates (identified by entity_id) who have been incarcerated in the last 5 years and have shown signs of mental health issues. The label 1 is assigned if they have been booked between the 'as_of_date' and six months forward, and if their incarceration duration exceeds two weeks. The label 0 is assigned otherwise. This process ensures the focus is on those likely to face more severe recidivism, aligning with our project's goal of targeting significant cases for intervention. The query incorporates multiple data sources, including jail records, mental health assessments, and emergency medical data, to comprehensively assess each individual's risk."

Train-validation splits:

Below are the 5 train-validation splits used in our model training and validation.

Train Start	Train End	Test Start	Test End
2013-01-01, 2013-07-01, 2014-01-01, 2014-07-01, 2015-01-01, 2015-07-01, 2016-01-01, 2016-07-01, 2017-01-01, 2017-07-01, 2018-01-01	Add 6 months to each	2018-07-01	2019-01-01
2011-01-01, 2011-07-01, 2012-01-01, 2012-07-01, 2013-01-01, 2013-07-01, 2014-01-01, 2014-07-01, 2015-01-01, 2015-07-01, 2016-01-01	Add 6 months to each	2016-07-01	2017-01-01
2010-01-01, 2010-07-01, 2011-01-01, 2011-07-01, 2012-01-01, 2012-07-01, 2013-01-01, 2013-07-01, 2014-01-01	Add 6 months to each	2014-07-01	2015-01-01
2010-01-01, 2010-07-01, 2011-01-01, 2011-07-01, 2012-01-01	Add 6 months to each	2012-07-01	2013-01-01
2010-01-01	2010-07-01	2010-07-01	2011-01-01

The temporal graph of precision@100 for each validation set for all the models in the grid (line color by model type)



Criteria used to select top models:

average precision@_100_abs over time

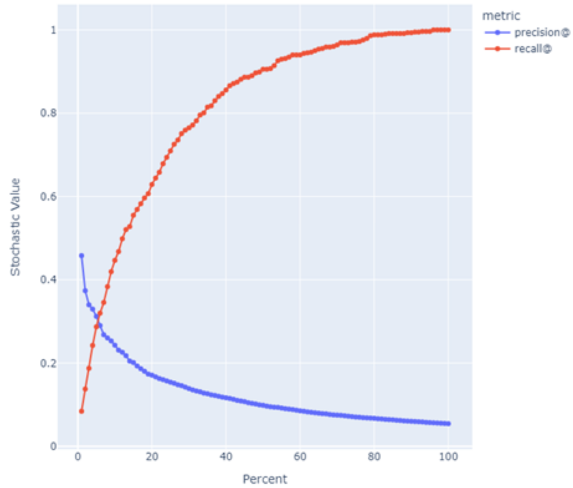
Top 5 models and the best baseline

Model	Model Group ID	HyperParameters	Average Precision @ 100
AdaBoost	20	{"algorithm": "SAMME.R", "n_estimators": 300, "learning_rate": 0.1}	**0.412**
AdaBoost	13	{"algorithm": "SAMME", "n_estimators": 1000, "learning_rate": 0.1}	**0.394**
AdaBoost	9	{"algorithm": "SAMME", "n_estimators": 1000, "learning_rate": 1}	**0.392**
Scaled Logistic Regression	32	{"C": 1.0, "solver": "saga", "penalty": "l1"}	**0.392**
Adaboost	8	{"algorithm": "SAMME", "n_estimators": 300, "learning_rate": 1}	**0.383**
Baseline: **Rank Booking Recency**	37	{"rules":[{"feature": "booking_entity_id_all_days_since_bookings_min", "low_value_high_score": true}, {"feature": "bjmhs2_entity_id_all_days_since_bjmhs_min", "low_value_high_score": true}]}	**0.253**

PR-k graphs

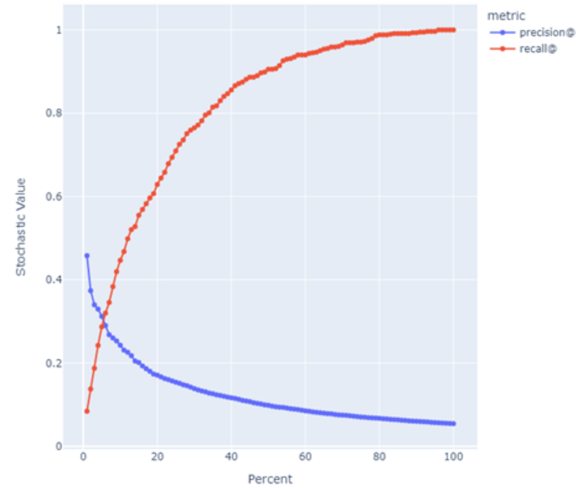
Model Group 20

AdaBoostClassifier
"algorithm": "SAMME.R", "learning_rate": 0.1, "n_estimators": 300



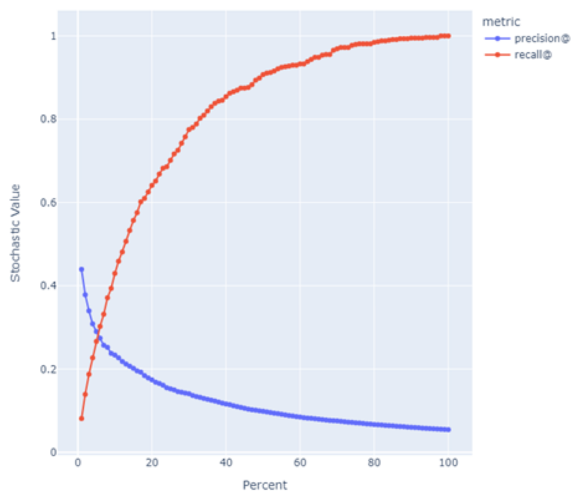
Model Group 13

AdaBoostClassifier
"algorithm": "SAMME.R", "learning_rate": 0.1, "n_estimators": 300



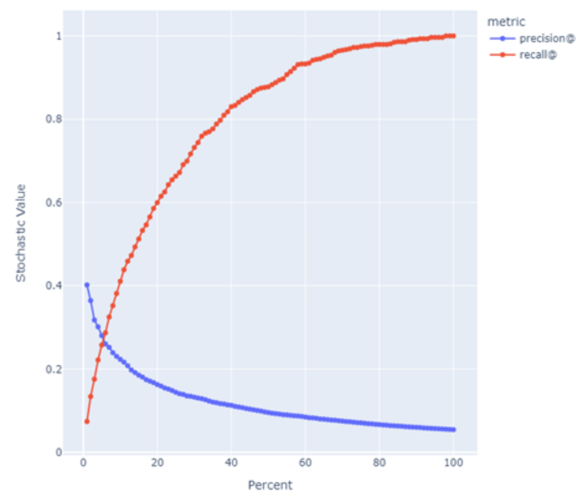
Model Group 9

AdaBoostClassifier
"algorithm": "SAMME", "learning_rate": 1, "n_estimators": 1000

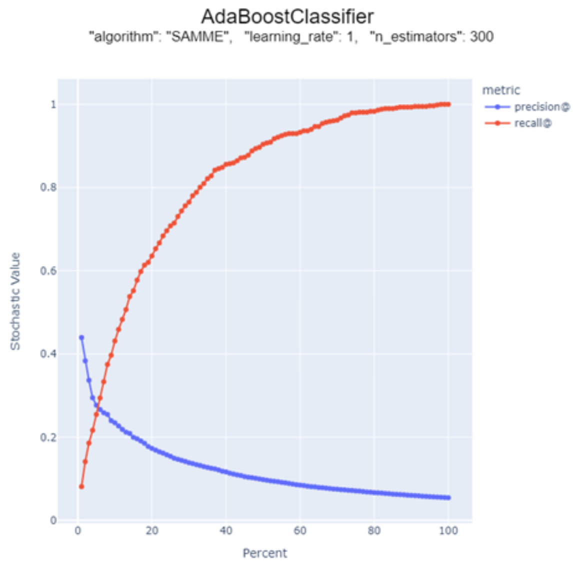


Model Group 32

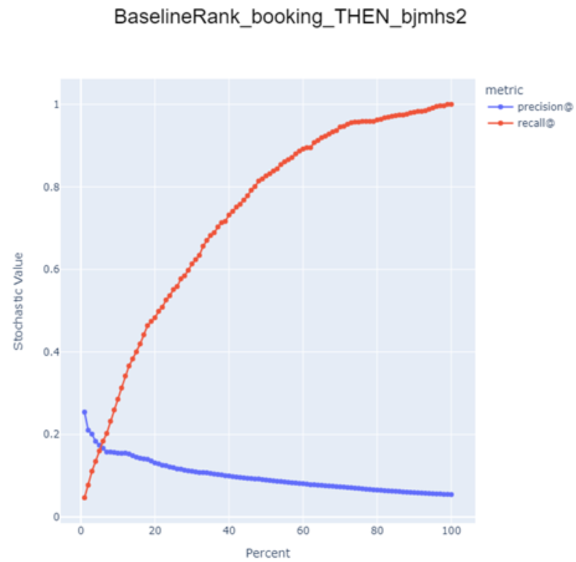
ScaledLogisticRegression
"C": 1.0, "penalty": "l1", "solver": "saga"



Model Group 8



Model Group 37



Feature importances for all features

Model Group 20, AdaBoost {"algorithm": "SAMME.R", "n_estimators": 300, "learning_rate": 0.1}

Feature	Value
all_days_since_bookings_min	0.1
all_avg_incarc_days_avg	0.093
all_days_since_lsir_min	0.077
all_prior_count	0.05
all_age_at_bkg_max	0.05
all_current_age_max	0.04
1 year_dispo_B_sum	0.033
5 years_avg_incarc_days_avg	0.03
6 months_dispo_B_sum	0.03
3 years_dispo_SS_sum	0.03
1 year_dispo_SI_sum	0.027
all_days_since_jcmhc_call_max	0.023
all_sex_MALE_max	0.023
1 year_avg_incarc_days_avg	0.02
5 years_dispo_PG_sum	0.02
3 years_dispo_B_sum	0.02
1 year_dispo_SS_sum	0.02
all_days_since_lsir_imp	0.02
3 years_dispo_SI_sum	0.02
6 months_arr_type_S_sum	0.017
5 years_dispo_RO_sum	0.017
3 years_prior_count	0.017
1 year_mhc_income_imp	0.013
3 years_dispo_WR_sum	0.013
5 years_days_next_bkg_avg	0.013
5 years_charge_type__NULL_sum	0.013
5 years_charge_type_C_sum	0.01
all_race_B_max	0.01
all_days_since_lsir_max	0.01
1 year_dispo_WR_sum	0.01
3 years_avg_incarc_days_avg	0.01
6 months_dispo_SI_sum	0.007
1 year_charge_type_F_sum	0.007
1 year_days_next_bkg_avg	0.007
3 years_dispo_PG_sum	0.007
1 year_mhc_income_avg	0.007
all_prior_count	0.007
5 years_charge_type_F_sum	0.007
6 months_dispo__NULL_sum	0.007
3 years_arr_type_S_sum	0.007
3 years_charge_type_F_sum	0.007

1 year_arr_type_S_sum	0.007
5 years_dispo_B_sum	0.007
1 year_charge_type_C_sum	0.003
1 year_prior_count	0.003
1 year_mhc_voc_status_no_activ_max	0.003
1 year_avg_incarc_days_imp	0.003
5 years_arr_type_S_sum	0.003
6 months_charge_type_F_sum	0.003
1 year_pta_employed_no_max	0.003
5 years_prior_count	0.003
all_days_since_bjmhs_min	0.003
all_days_since_medact_min	0.003
1 year_charge_type__NULL_sum	0.003
3 years_charge_type__NULL_sum	0.003

Model Group 13, AdaBoost {"algorithm": "SAMME", "n_estimators": 1000, "learning_rate": 0.1}

Feature	Value
1 year_days_next_bkg_avg	0.268
all_days_since_bookings_min	0.201
1 year_avg_incarc_days_imp	0.065
all_avg_incarc_days_avg	0.056
3 years_avg_incarc_days_avg	0.034
1 year_dispo_B_sum	0.031
all_age_at_bkg_max	0.03
all_current_age_max	0.026
6 months_dispo_B_sum	0.025
1 year_dispo_SI_sum	0.02
3 years_dispo_B_sum	0.019
all_days_since_lsir_min	0.018
6 months_arr_type_S_sum	0.017
6 months_dispo__NULL_sum	0.015
all_prior_count	0.014
1 year_arr_type_S_sum	0.011
1 year_avg_incarc_days_avg	0.009
3 years_prior_count	0.009
all_days_since_lsir_imp	0.008
all_days_since_jcmhc_call_max	0.008
5 years_dispo_B_sum	0.008
5 years_dispo_PG_sum	0.008
5 years_avg_incarc_days_avg	0.007
all_sex_MALE_max	0.007
1 year_dispo_SS_sum	0.007
5 years_dispo_RO_sum	0.007
3 years_dispo_SS_sum	0.007

3 years_dispo_WR_sum	0.006
3 years_dispo_PG_sum	0.006
1 year_mhc_income_imp	0.006
3 years_dispo_SI_sum	0.005
5 years_days_next_bkg_avg	0.004
5 years_charge_type__NULL_sum	0.004
6 months_dispo_SI_sum	0.003
3 years_arr_type_S_sum	0.003
3 years_charge_type__NULL_sum	0.003
5 years_prior_count	0.003
1 year_dispo_WR_sum	0.003
1 year_prior_count	0.002
all_days_since_lsir_max	0.002
1 year_charge_type_F_sum	0.002
6 months_charge_type_F_sum	0.002
5 years_arr_type_S_sum	0.002
all_race_B_max	0.002
1 year_charge_type__NULL_sum	0.001
5 years_charge_type_C_sum	0.001
1 year_mhc_income_avg	0.001
3 years_charge_type_F_sum	0.001
5 years_charge_type_F_sum	0.001
1 year_mhc_voc_status_no_activ_max	0.001
1 year_charge_type_C_sum	0.001

Model Group 9, AdaBoost {"algorithm": "SAMME", "n_estimators": 1000, "learning_rate": 1}

Feature	Value
1 year_days_next_bkg_avg	0.202
1 year_avg_incarc_days_imp	0.1
all_days_since_bookings_min	0.069
all_age_at_bkg_max	0.066
3 years_avg_incarc_days_avg	0.049
1 year_avg_incarc_days_avg	0.043
all_avg_incarc_days_avg	0.041
all_current_age_max	0.037
5 years_dispo_B_sum	0.037
all_days_since_lsir_min	0.031
6 months_dispo_B_sum	0.026
all_prior_count	0.024
3 years_prior_count	0.02
1 year_pta_employed_no_max	0.015
all_prior_count	0.014
5 years_charge_type_C_sum	0.012

1 year_dispo_B_sum	0.012
all_days_since_bookings_max	0.012
1 year_dispo_SI_sum	0.012
all_days_since_lsir_imp	0.012
3 years_dispo_SS_sum	0.01
6 months_charge_type_F_sum	0.01
5 years_avg_incarc_days_avg	0.01
3 years_dispo_SI_sum	0.009
3 years_dispo_B_sum	0.008
3 years_dispo_PG_sum	0.008
5 years_dispo_PG_sum	0.008
6 months_arr_type_S_sum	0.008
1 year_arr_type_S_sum	0.008
1 year_charge_type__NULL_sum	0.007
all_sex_MALE_max	0.007
all_days_since_jcmhc_call_min	0.006
5 years_dispo_RO_sum	0.006
1 year_charge_type_F_sum	0.006
5 years_days_next_bkg_avg	0.006
3 years_dispo_WR_sum	0.005
6 months_dispo_SI_sum	0.005
1 year_mhc_income_imp	0.005
all_race_B_max	0.004
5 years_dispo_SS_sum	0.004
6 months_mhc_voc_status_home_care_max	0.003
3 years_days_next_bkg_avg	0.003
6 months_dispo_RO_sum	0.003
1 year_mhc_voc_status_no_activ_max	0.002
5 years_arr_type_S_sum	0.002
1 year_mhc_income_avg	0.002
all_days_since_lsir_max	0.002
1 year_dispo_WR_sum	0.002
all_days_since_bjmhs_min	0.002
1 year_prior_count	0.002
all_days_since_medact_min	0.002
6 months_mhc_income_avg	0.001

Model Group 32, Scaled Logistic Regression {"C": 1.0, "solver": "saga", "penalty": "l1"}

Feature	**Value**
all_avg_incarc_days_avg	47.787
all_prior_count	7.078
6 months_dispo_B_sum	5.317
1 year_dispo_B_sum	4.298

5 years_charge_type_X_sum	2.857
1 year_arr_type_S_sum	2.726
3 years_charge_type__NULL_sum	2.718
1 year_charge_type_F_sum	2.643
all_days_since_bjmhs_max	2.617
5 years_dispo_SI_sum	2.421
5 years_arr_type_R_sum	2.306
3 years_dispo_PG_sum	2.197
5 years_dispo_PG_sum	2.058
5 years_avg_incarc_days_avg	2.049
6 months_charge_type_F_sum	2.048
6 months_ mhc_voc_status_home_care_max	1.922
3 years_days_next_bk_imp	1.864
6 months_arr_type_S_sum	1.74
1 year_charge_type__NULL_sum	1.731
5 years_charge_type_F_sum	1.716
1 year_dispo_RO_sum	1.706
3 years_avg_incarc_days_avg	1.69
all_prior_count	1.675
3 years_prior_count	1.65

1 year_prior_count	1.594
all_days_next_bk_imp	1.587
5 years_days_next_bk_imp	1.587
3 years_avg_incarc_days_imp	1.513
1 year_charge_type_M_sum	1.507
3 years_arr_type_T_sum	1.419
1 year_dispo_PG_sum	1.397
1 year_pta_employed__NULL_max	1.395
1 year_bjmhs_max	1.394
6 months_ mhc_voc_status_prevoc_max	1.384
5 years_dispo_RO_sum	1.366
3 years_dispo_B_sum	1.309
5 years_charge_type_ON_sum	1.287
5 years_arr_type_S_sum	1.268
all_days_since_pta_min	1.258
all_sex_MALE_max	1.257
all_days_since_pta_max	1.245
all_days_since_pta_imp	1.239
all_days_since_pt_imp	1.239
6 months_ mhc_voc_status_active_max	1.23

3 years_charge_type_X_sum	1.218
3 years_charge_type_F_sum	1.215
6 months_arr_type_O_sum	1.207
1 year_arr_type__NULL_sum	1.204
3 years_arr_type_R_sum	1.2
6 months_ mhc_voc_status__NULL_max	1.199
6 months_bjmhs_max	1.193
1 year_ mhc_voc_status_no_activ_max	1.15
1 year_joco_resident__NULL_max	1.135
1 year_marital_status__NULL_max	1.135
1 year_top20_zip_imp	1.135
all_days_since_jcmhc_call_max	1.128
1 year_top20_zip_max	1.075
all_race_B_max	1.068
5 years_arr_type__NULL_sum	1.048
3 years_arr_type_S_sum	1.048
1 year_marital_status_W_max	1.041
4 years_bjmhs_imp	1.04
all_bjmhs_imp	1.04

5 years_bjmhs_imp	1.04
all_days_since_bookings_max	1.04
all_days_since_jcmhc_adm_min	1.04
6 months_bjmhs_imp	1.035
1 year_ mhc_voc_status_job_30_less_max	1.034
1 year_arr_type_O_sum	1.027
1 year_marital_status_S_max	1.021
all_race_I_max	1.021
1 year_marital_status_D_max	1.013
all_substance_flag_pos_sum	1.0
1 year_arr_type_R_sum	1.0
1 year_arr_type_T_sum	1.0
3 years_arr_type_O_sum	1.0
5 years_prior_count	1.0
6 months_arr_type_R_sum	1.0
6 months_prior_count	1.0
all_mh_flag_count	1.0
all_days_since_bjmhs_imp	1.0
2 years_bjmhs_imp	1.0
3 years_bjmhs_imp	1.0
1 year_charge_type_FU_sum	1.0

1 year_charge_type_ON_sum	1.0
1 year_charge_type_OY_sum	1.0
1 year_charge_type_T_sum	1.0
1 year_dispo_D_sum	1.0
1 year_dispo_HA_sum	1.0
1 year_dispo_P_sum	1.0
1 year_dispo_R_sum	1.0
1 year_dispo_S_sum	1.0
1 year_dispo_TR_sum	1.0
1 year_dispo_T_sum	1.0
3 years_charge_type_C_sum	1.0
3 years_charge_type_FU_sum	1.0
3 years_charge_type_M_sum	1.0
3 years_charge_type_ON_sum	1.0
3 years_charge_type_OY_sum	1.0
3 years_charge_type_T_sum	1.0
3 years_dispo_HA_sum	1.0
3 years_dispo_P_sum	1.0
3 years_dispo_RO_sum	1.0
3 years_dispo_R_sum	1.0
3 years_dispo_SS_sum	1.0
3 years_dispo_S_sum	1.0
3 years_dispo_TR_sum	1.0

3 years_dispo_T_sum	1.0
5 years_charge_type_FU_sum	1.0
5 years_charge_type__NULL_sum	1.0
5 years_charge_type_OY_sum	1.0
5 years_charge_type_T_sum	1.0
5 years_dispo_D_sum	1.0
5 years_dispo__NULL_sum	1.0
5 years_dispo_P_sum	1.0
5 years_dispo_R_sum	1.0
5 years_dispo_S_sum	1.0
5 years_dispo_TR_sum	1.0
5 years_dispo_T_sum	1.0
5 years_dispo_WR_sum	1.0
6 months_charge_type_C_sum	1.0
6 months_charge_type_FU_sum	1.0
6 months_charge_type_M_sum	1.0
6 months_ charge_type__NULL_sum	1.0
6 months_charge_type_OY_sum	1.0
6 months_charge_type_T_sum	1.0
6 months_charge_type_X_sum	1.0

6 months_dispo_HA_sum	1.0
6 months_dispo_P_sum	1.0
6 months_dispo_RO_sum	1.0
6 months_dispo_R_sum	1.0
6 months_dispo_S_sum	1.0
6 months_dispo_TR_sum	1.0
6 months_dispo_T_sum	1.0
all_current_age_max	1.0
all_race__NULL_max	1.0
all_race_O_max	1.0
all_race_W_max	1.0
all_sex_FEMALE_max	1.0
all_sex__NULL_max	1.0
1 year_marital_status_OTHER_max	1.0
1 year_pta_employed_yes_max	1.0
1 year_mhc_income_avg	1.0
6 months_mhc_income_avg	1.0
1 year_ mhc_voc_status_home_care_max	1.0
1 year_ mhc_voc_status_job_30_more_max	1.0

1 year_mhc_voc_status_other_max	1.0
1 year_ mhc_voc_status_volunteer_max	1.0
6 months_ mhc_voc_status_no_activ_max	1.0
6 months_ mhc_voc_status_other_max	1.0
6 months_ mhc_voc_status_retired_max	1.0
6 months_ mhc_voc_status_volunteer_max	1.0
all_drug_flag_neg_sum	1.0
all_drug_flag_pos_sum	1.0
1 year_days_next_bk_imp	1.0
all_days_since_jcdhe_imp	1.0
all_days_since_jcdhe_max	1.0
all_days_since_jcdhe_min	1.0
all_days_since_medact_max	1.0
all_prior_count	1.0
all_prior_imp	1.0
all_risk_level_High_count	1.0

all_risk_level_Low_count	1.0
all_risk_level_Moderate_count	1.0
all_risk_level__NULL_count	1.0
all_substance_flag_neg_sum	1.0
1 year_ mhc_voc_status_retired_max	0.994
1 year_marital_status_M_max	0.988
6 months_prior_imp	0.985
1 year_prior_imp	0.985
5 years_prior_imp	0.985
3 years_prior_imp	0.985
5 years_dispo_HA_sum	0.985
all_days_since_jcmhc_ad_imp	0.972
all_prior_imp	0.972
5 years_arr_type_O_sum	0.971
1 year_joco_resident_false_max	0.971
5 years_charge_type_M_sum	0.97
5 years_dispo_SS_sum	0.968
all_days_since_jcmhc_adm_max	0.966
3 years_dispo__NULL_sum	0.944
all_drug_flag_po_imp	0.917
all_substance_flag_neg_imp	0.917

all_mh_flag_imp	0.917
all_substance_flag_po_imp	0.917
all_drug_flag_neg_imp	0.917
4 years_bjmhs_max	0.909
2 years_bjmhs_max	0.909
3 years_bjmhs_max	0.909
all_bjmhs_max	0.909
5 years_bjmhs_max	0.909
3 years_dispo_WR_sum	0.902
1 year_mhc_voc_status_active_max	0.899
1 year_ mhc_voc_status__NULL_max	0.891
6 months_mhc_income_imp	0.869
1 year_dispo__NULL_sum	0.868
1 year_bjmhs_imp	0.862
6 months_dispo__NULL_sum	0.861
3 years_arr_type__NULL_sum	0.859
1 year_mhc_income_imp	0.857
all_days_since_lsir_max	0.856
6 months_ mhc_voc_status_job_30_less_max	0.844

6 months_ mhc_voc_status_job_30_more _max	0.832
1 year_joco_resident_true_max	0.825
1 year_avg_incarc_days_imp	0.824
1 year_charge_type_C_sum	0.823
1 year_days_next_bkg_avg	0.819
5 years_dispo_B_sum	0.803
1 year_ mhc_voc_status_prevoc_max	0.802
3 years_dispo_D_sum	0.8
6 months_arr_type__NULL_sum	0.789
all_race_A_max	0.765
all_days_since_medact_imp	0.746
6 months_arr_type_T_sum	0.744
1 year_avg_incarc_days_avg	0.736
6 months_charge_type_ON_sum	0.726
all_days_since_jcmhc_call_imp	0.7
3 years_days_next_bkg_avg	0.699
all_days_since_jcmhc_call_min	0.673
all_days_next_bkg_avg	0.666
6 months_dispo_WR_sum	0.558
all_days_since_lsir_imp	0.549

1 year_dispo_SS_sum	0.547
all_days_since_bjmhs_min	0.512
5 years_arr_type_T_sum	0.512
6 months_dispo_PG_sum	0.486
1 year_pta_employed_no_max	0.483
1 year_dispo_WR_sum	0.433
all_days_since_medact_min	0.405
all_days_since_lsir_min	0.397
6 months_dispo_SS_sum	0.397
5 years_days_next_bkg_avg	0.367
all_age_at_bkg_max	0.303
6 months_dispo_D_sum	0.243
1 year_charge_type_X_sum	0.234
5 years_charge_type_C_sum	0.21
6 months_dispo_SI_sum	0.144
3 years_dispo_SI_sum	0.132
1 year_dispo_SI_sum	0.091
all_days_since_bookings_min	0.022

Model Group 8, AdaBoost {"algorithm": "SAMME", "n_estimators": 300, "learning_rate": 1}

Feature	Value
1 year_days_next_bkg_avg	0.236
1 year_avg_incarc_days_imp	0.114
all_days_since_bookings_min	0.067
3 years_avg_incarc_days_avg	0.057
all_avg_incarc_days_avg	0.046
5 years_dispo_B_sum	0.043
all_days_since_lsir_min	0.034
all_age_at_bkg_max	0.033
6 months_dispo_B_sum	0.03
all_current_age_max	0.025
1 year_avg_incarc_days_avg	0.024
3 years_prior_count	0.021
all_prior_count	0.021
all_prior_count	0.016
all_days_since_lsir_imp	0.014
1 year_dispo_B_sum	0.013
5 years_charge_type_C_sum	0.013
3 years_dispo_SS_sum	0.012
6 months_charge_type_F_sum	0.012
1 year_dispo_SI_sum	0.011
3 years_dispo_SI_sum	0.01
3 years_dispo_B_sum	0.01
5 years_avg_incarc_days_avg	0.009
3 years_dispo_PG_sum	0.009
5 years_dispo_PG_sum	0.009
6 months_arr_type_S_sum	0.009
1 year_arr_type_S_sum	0.009
1 year_charge_type__NULL_sum	0.008
all_days_since_jcmhc_call_min	0.007
5 years_dispo_RO_sum	0.007
1 year_charge_type_F_sum	0.007
5 years_days_next_bkg_avg	0.007
6 months_dispo_SI_sum	0.006
all_days_since_bookings_max	0.006
1 year_mhc_income_imp	0.005
3 years_dispo_WR_sum	0.005
all_sex_MALE_max	0.005
all_race_B_max	0.005
5 years_dispo_SS_sum	0.005
1 year_pta_employed_no_max	0.004
3 years_days_next_bkg_avg	0.003
6 months_dispo_RO_sum	0.003

1 year_mhc_voc_status_no_activ_max	0.003
1 year_mhc_income_avg	0.002
1 year_prior_count	0.002
all_days_since_bjmhs_min	0.001

Model Group 37, Baseline Model {"rules":[{"feature": "booking_entity_id_all_days_since_bookings_min", "low_value_high_score": true}, {"feature": "bjmhs2_entity_id_all_days_since_bjmhs_min", "low_value_high_score": true}]}

Feature	Value
all_days_since_bjmhs_min	1.0
all_days_since_bookings_min	1.0

Cross-tabs for 10 most different features

Model Group 20, AdaBoost {"algorithm": "SAMME.R", "n_estimators": 300, "learning_rate": 0.1}

	feature	feature_importance	val_top_k	val_bottom_k	ratio
0	6 months_dispo_SI_sum	0.007	0.000	0.010	1.000000
1	1 year_avg_incarc_days_imp	0.003	0.000	0.563	1.000000
2	1 year_dispo_SI_sum	0.027	0.000	0.028	1.000000
3	6 months_dispo__NULL_sum	0.007	0.000	0.771	1.000000
4	1 year_days_next_bkg_avg	0.007	187.413	17787.326	0.989464
5	5 years_days_next_bkg_avg	0.013	223.795	7343.130	0.969523
6	6 months_arr_type_S_sum	0.017	0.220	0.017	0.922727
7	1 year_charge_type_C_sum	0.003	0.160	0.013	0.918750
8	6 months_charge_type_F_sum	0.003	0.970	0.079	0.918557
9	6 months_dispo_B_sum	0.030	1.990	0.182	0.908543

Model Group 13, AdaBoost {"algorithm": "SAMME", "n_estimators": 1000, "learning_rate": 0.1}

	feature	feature_importance	val_top_k	val_bottom_k	ratio
0	1 year_dispo_SI_sum	0.020	0.000	0.028	1.000000
1	1 year_avg_incarc_days_imp	0.065	0.000	0.563	1.000000
2	6 months_dispo_SI_sum	0.003	0.000	0.010	1.000000
3	6 months_dispo__NULL_sum	0.015	0.000	0.771	1.000000
4	1 year_days_next_bkg_avg	0.268	186.216	17787.337	0.989531
5	5 years_days_next_bkg_avg	0.004	230.151	7343.070	0.968657
6	1 year_charge_type_C_sum	0.001	0.180	0.013	0.927778
7	6 months_arr_type_S_sum	0.017	0.220	0.017	0.922727
8	6 months_charge_type_F_sum	0.002	0.970	0.079	0.918557
9	6 months_dispo_B_sum	0.025	1.870	0.183	0.902139

Model Group 9, AdaBoost {"algorithm": "SAMME", "n_estimators": 1000, "learning_rate": 1}

	feature	feature_importance	val_top_k	val_bottom_k	ratio
0	1 year_avg_incarc_days_imp	0.100	0.000	0.563	1.000000
1	6 months_mhc_voc_status_home_care_max	0.003	0.000	0.001	1.000000
2	1 year_days_next_bkg_avg	0.202	216.804	17787.049	0.987811
3	3 years_days_next_bkg_avg	0.003	266.793	11283.361	0.976355
4	5 years_days_next_bkg_avg	0.006	252.027	7342.864	0.965677
5	6 months_charge_type_F_sum	0.010	1.100	0.078	0.929091
6	6 months_arr_type_S_sum	0.008	0.200	0.017	0.915000
7	6 months_dispo_B_sum	0.026	1.840	0.183	0.900543
8	6 months_dispo_RO_sum	0.003	0.470	0.047	0.900000
9	1 year_arr_type_S_sum	0.008	0.330	0.036	0.890909

Model Group 32, Scaled Logistic Regression {"C": 1.0, "solver": "saga", "penalty": "l1"}

	feature	feature_importance	val_top_k	val_bottom_k	ratio
0	1 year_dispo__NULL_sum	0.868	0.0	0.583	1.0
1	all_days_next_bk_imp	1.587	0.0	0.271	1.0
2	3 years_avg_incarc_days_imp	1.513	0.0	0.211	1.0
3	3 years_dispo__NULL_sum	0.944	0.0	0.220	1.0
4	1 year_mhc_voc_status_retired_max	0.994	0.0	0.001	1.0
5	all_days_since_jcdhe_min	1.000	0.0	0.045	1.0
6	all_days_since_jcdhe_max	1.000	0.0	0.045	1.0
7	1 year_days_next_bk_imp	1.000	0.0	0.703	1.0
8	all_drug_flag_pos_sum	1.000	0.0	0.002	1.0
9	6 months_mhc_voc_status_retired_max	1.000	0.0	0.001	1.0

Model Group 8, AdaBoost {"algorithm": "SAMME", "n_estimators": 300, "learning_rate": 1}

	feature	feature_importance	val_top_k	val_bottom_k	ratio
0	1 year_avg_incarc_days_imp	0.114	0.000	0.563	1.000000
1	1 year_days_next_bkg_avg	0.236	218.096	17787.037	0.987738
2	3 years_days_next_bkg_avg	0.003	266.734	11283.362	0.976360
3	5 years_days_next_bkg_avg	0.007	251.323	7342.871	0.965773
4	6 months_charge_type_F_sum	0.012	1.130	0.078	0.930973
5	6 months_arr_type_S_sum	0.009	0.200	0.017	0.915000
6	6 months_dispo_B_sum	0.030	1.840	0.183	0.900543
7	6 months_dispo_RO_sum	0.003	0.460	0.047	0.897826
8	all_days_since_bookings_min	0.067	70.170	633.710	0.889271
9	1 year_arr_type_S_sum	0.009	0.300	0.036	0.880000

Model Group 37, Baseline Model {"rules":[{"feature": "booking_entity_id_all_days_since_bookings_min", "low_value_high_score": true}, {"feature": "bjmhs2_entity_id_all_days_since_bjmhs_min", "low_value_high_score": true}]}

	feature	feature_importance	val_top_k	val_bottom_k	ratio
0	all_days_since_bookings_min	1.0	5.90	634.316	0.990699
1	all_days_since_bjmhs_min	1.0	18.89	114.203	0.834593

Bias metrics

Model Group 20, AdaBoost {"algorithm": "SAMME.R", "n_estimators": 300, "learning_rate": 0.1}

attribute_value	total_entities	group_label_pos	group_label_neg	group_size	tp	tpr	tpr_disparity	tpr_ref_group_value
FEMALE	10716	153	3436	3589	10	0.065	0.826	MALE
low_income	10716	496	8869	9365	36	0.073	0.767	not_low_income
BLACK	10716	134	1866	2000	11	0.082	1.118	WHITE

Model Group 13, AdaBoost {"algorithm": "SAMME", "n_estimators": 1000, "learning_rate": 0.1}

attribute_value	total_entities	group_label_pos	group_label_neg	group_size	tp	tpr	tpr_disparity	tpr_ref_group_value
FEMALE	10716	153	3436	3589	9	0.059	0.791	MALE
low_income	10716	496	8869	9365	33	0.067	0.703	not_low_income
BLACK	10716	134	1866	2000	9	0.067	0.945	WHITE

Model Group 9, AdaBoost {"algorithm": "SAMME", "n_estimators": 1000, "learning_rate": 1}

attribute_value	total_entities	group_label_pos	group_label_neg	group_size	tp	tpr	tpr_disparity	tpr_ref_group_value
FEMALE	10716	153	3436	3589	9	0.059	0.767	MALE
low_income	10716	496	8869	9365	36	0.073	1.074	not_low_income
BLACK	10716	134	1866	2000	13	0.097	1.511	WHITE

Model Group 32, Scaled Logistic Regression {"C": 1.0, "solver": "saga", "penalty": "l1"}

attribute_value	total_entities	group_label_pos	group_label_neg	group_size	tp	tpr	tpr_disparity	tpr_ref_group_value
FEMALE	10716	153	3436	3589	11	0.072	1.034	MALE
low_income	10716	496	8869	9365	34	0.069	0.845	not_low_income
BLACK	10716	134	1866	2000	9	0.067	0.945	WHITE

Model Group 8, AdaBoost {"algorithm": "SAMME", "n_estimators": 300, "learning_rate": 1}

attribute_value	total_entities	group_label_pos	group_label_neg	group_size	tp	tpr	tpr_disparity	tpr_ref_group_value
FEMALE	10716	153	3436	3589	12	0.078	1.090	MALE
low_income	10716	496	8869	9365	36	0.073	0.895	not_low_income
BLACK	10716	134	1866	2000	14	0.104	1.627	WHITE

Model Group 37, Baseline Model {"rules":[{"feature": "booking_entity_id_all_days_since_bookings_min", "low_value_high_score": true}, {"feature": "bjmhs2_entity_id_all_days_since_bjmhs_min", "low_value_high_score": true}]}

attribute_value	total_entities	group_label_pos	group_label_neg	group_size	tp	tpr	tpr_disparity	tpr_ref_group_value
FEMALE	10716	153	3436	3589	8	0.052	1.363	MALE
low_income	10716	496	8869	9365	19	0.038	0.567	not_low_income
BLACK	10716	134	1866	2000	7	0.052	1.340	WHITE

References:

1. Bauman, Matthew J., Kate S. Boxer, Tzu-Yun Lin, Erika Salomon, Hareem Naveed, Lauren Haynes, Joe Walsh, et al. “Reducing Incarceration through Prioritized Interventions.” In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, 1–8. COMPASS ’18. New York, NY, USA: Association for Computing Machinery, 2018. https://doi.org/10.1145/3209811.3209869.