

Reducing Incarceration through Prioritized Interventions

Erika Salomon
University of Chicago
ecsalomon@uchicago.edu

Matthew J. Bauman
University of Chicago
mbauman@uchicago.edu

Tzu-Yun Lin
University of Chicago
tlin2@uchicago.edu

Kate Boxer
Columbia University
ksbox12@gmail.com

Hareem Naveed
University of Chicago
hareem@uchicago.edu

Lauren Haynes
University of Chicago
lnhaynes@uchicago.edu

Joe Walsh
University of Chicago
jtwalsh@uchicago.edu

Jen Helsby
University of Chicago
jen@redshiftzero.com

Steve Yoder
Johnson County, KS
Steve.Yoder@jocogov.org

Robert Sullivan
Johnson County, KS
Robert.Sullivan@jocogov.org

Chris Schneweis
Johnson County, KS
Chris.Schneweis@jocogov.org

Rayid Ghani
University of Chicago
rayid@uchicago.edu

ABSTRACT

The most vulnerable individuals in society often struggle with long-lasting, multi-faceted challenges like mental illness, substance abuse, chronic health conditions, and homelessness. Individuals experiencing these difficulties tend to interact with public services and departments frequently, but many communities are struggling to identify those individuals, let alone meet their needs in meaningful and cost-effective ways. In this paper, we describe our work with Johnson County, Kansas, that uses machine learning to prioritize outreach to individuals most at risk of being booked into jail within the next year. For the first time, we brought together Johnson County's jail, emergency medical, and mental health data, identified individuals who touch multiple systems, and built a model to predict individual jail bookings. Our system significantly outperformed both a random baseline and several simple heuristics that domain experts are likely to use and implement. By focusing on 200 individuals (which is the intervention capacity of Johnson County) who had interacted with both mental health services and the criminal justice system, we predicted jail bookings in the following year with 51% accuracy. This work provides a framework and prototype system for Johnson County as well as many other jurisdictions that are part of the Data Driven Justice Initiative as they develop intervention models to proactively connect social and mental health workers with individuals in need of care to avoid incarceration.

ACM Reference format:

Erika Salomon, Matthew J. Bauman, Tzu-Yun Lin, Kate Boxer, Hareem Naveed, Lauren Haynes, Joe Walsh, Jen Helsby, Steve Yoder, Robert Sullivan, Chris Schneweis, and Rayid Ghani. 2017. Reducing Incarceration through

This work was done as part of the 2016 Eric and Wendy Schmidt Data Science for Social Good Fellowship at the University of Chicago.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGKDD'17, Halifax, Nova Scotia, Canada

© 2017 Copyright held by the owner/author(s). 978-x-xxxx-xxxx-x/YY/MM...\$15.00
DOI: 10.1145/nnnnnnn.nnnnnnn

Prioritized Interventions. In *Proceedings of ACM SIGKDD, Halifax, Nova Scotia, Canada, August 2017 (SIGKDD'17)*, 7 pages.

DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

With millions of people moving through jails every year in the US, the criminal justice system is larger than ever. However, many bookings are a result of low-level, non-violent misdemeanors, costing local governments billions of dollars a year. In local jails, 64% of people struggle with mental illness, 68% have a substance abuse disorder, and 44% suffer from chronic health issues [4]. Communities across the country have recognized that a relatively small number of these highly vulnerable people cycle repeatedly not just through local jails, but also emergency medical services (EMS), hospital emergency rooms, shelters, and other public systems, receiving fragmented and uncoordinated care with poor outcomes at great cost. With three times more people with mental health problems in jails or prison than there are in hospitals [11], local police, emergency medical teams, and jails have become the front lines for people with complex social and behavioral health issues.

Local governments, policymakers, and practitioners seek to break this kind of cycle of incarceration through policy changes and early interventions that result in a more rational system. These reforms would ideally not only help vulnerable individuals but also increase public safety and reduce money spent on incarceration. Many law enforcement officers, prosecutors, defenders, and judges believe too many people with mental illness become involved in the criminal justice system because the mental health system has failed them [9]. If, instead, people with mental illness received the services they needed in time, we could potentially prevent the situations where they would get arrested, go to jail, or face charges in court.

Unfortunately, in most jurisdictions—including Johnson County—mental health, EMS, and jail systems rarely share data that can help care providers understand and address their patient's complex needs. For example, mental health professionals may be lucky to learn that their patients have been booked into jail. This lack of

information not only makes it difficult to identify who is touching multiple systems but also to learn patterns that predict future system contacts. In particular, individuals with complex needs often end up in jail, which is more costly and less effective than direct treatment.

This is where our work comes in. By combining and modeling data from disparate systems, we can identify individuals at risk as early as possible. At that point simpler, more effective, and less costly interventions may directly address underlying issues before the criminal justice system ever gets involved.

In this paper, we describe a machine learning system that addresses this problem using data provided by several agencies running public systems in Johnson County, Kansas. Our predictive model performs nearly 500% better than random and 25% better than simple heuristics. Our system helps jurisdictions identify individuals with histories of mental illness and incarceration at risk of returning to jail. In doing so, it enables them to provide mental health and social service interventions to prevent jail time, improve the lives of residents, and enable jurisdictions to invest their resources more efficiently.

Johnson County is validating the lists of individuals our model deemed most at risk to determine appropriate intervention strategies. This has not only been a successful initiative for Johnson County in improving their outreach strategy to reduce incarceration; it can also be applied to other cities across the US. Our code has been released as open source and is available on github at <https://github.com/dssg/johnson-county-ddj-public> for other jurisdictions to reuse and extend. In addition, we are working with a consortium of over 130 jurisdictions as part of the Data Driven Justice Initiative to develop data integration and analytics infrastructure that can be used to proactively connect at-risk individuals with the appropriate social and health services.

2 CURRENT APPROACHES

Individuals with multi-faceted problems (e.g., homelessness, mental illness, and chronic health conditions) make repeated contacts with multiple service providers—including social and mental health services, emergency medical services, and hospitals—and with the criminal justice system. In many communities, these service providers struggle to offer coordinated care. To address the needs of this population, communities have implemented policy reforms intended to increase contact among service providers to meet the underlying needs of their residents. These efforts include criminal justice coordinating councils (bodies dedicated to innovations in local criminal justice that often include partners from social service and mental health organizations) and Sequential Intercept Model mapping [8] (a process that identifies what services are most appropriate for interventions at various stages in the criminal justice process). In addition, many service providers have begun engaging with frequent service users to understand and address their underlying needs, even when those needs are beyond the traditional purview of the service provider.

For example, in the medical community, patients who accumulate multiple emergency department visits and hospital admissions are the target of specific policy initiatives to prevent the use of

costly medical services by rerouting them to more effective primary care and community-based interventions [5]. Frequent visits to the emergency room are associated with a number of unmet underlying needs such as homelessness and mental illness [5]. Although such patients make up a small proportion of the population, their frequent visits make up a large proportion of costs to health-care providers and insurers [6]. Often, caring for these individuals' underlying needs may be less expensive than emergency room treatment while simultaneously improving the health and well-being of patients. Some hospitals have implemented programs to do exactly that. For example, the Better Health Through Housing at the University of Illinois at Chicago's hospital aims to improve health and reduce the cost of emergency care by providing housing and intensive case management to emergency room patients experiencing chronic homelessness. Such programs reduce costs while simultaneously improving care.

3 EXTENDING THE CURRENT SYSTEM

Our goal is to allow the criminal justice agencies to implement similar interventions to those used by hospitals, with the goal of integrating data from social services, mental health services, and criminal justice administration to identify individuals who needs bring them in repeated contact with a number of systems.

Like hospitals, local jails encounter repeated visits from individuals with underlying, unmet needs. Between 6%–15% of individuals in city and county jails have a severe mental illness [7]. Inmates with symptoms of mental illness are more likely to experience other vulnerabilities such as homelessness, unemployment, and trauma prior to incarceration [4]. Moreover, nearly 75% of jail detainees with a serious mental illness have a co-occurring substance abuse disorder [1]. Jail systems struggle to maintain adequate resources to provide effective treatment options, making them ill-suited to address the needs of this population. As a result, these inmates may undergo intensive supervision and reentry programs when they are released. By identifying these individuals early on, government agencies can address their underlying needs without incarceration, reduce criminal justice costs, and improve the well-being of the residents they serve.

The work described in this paper was initiated as part of the Obama White House's Data-Driven Justice Initiative. Johnson County, Kansas, is one of several partners around the country participating in the initiative to better understand residents' needs, while diverting them from the criminal justice system. Johnson County is the most populous county in Kansas, with approximately 575 000 residents. It consists of 20 municipalities and is governed by a Board of County Commissioners. Johnson County has already instituted some reforms designed to address the needs of individuals with multi-faceted problems. For example, mental-health social workers are embedded as co-responders in several police departments in the county in order to help prevent trips to emergency rooms and jail by providing alternate interventions and coordinating longer-term care. This program is being expanded to embed a social worker with the emergency medical services, as well. Working with Johnson County, we propose a machine learning system to identify and prioritize individuals and allow government agencies

and non-profits to provide targeted interventions to prevent recidivism and redirect people for more comprehensive care, rather than reactively dealing with them after they get booked in the criminal justice system.

4 OUR APPROACH

In this first iteration of the system, we have limited our analyses to individuals who have already interacted with both Johnson County's Mental Health Center, which provides all public mental health services for the county, and its jail system. We combined over six years of historical individual-level data from the county's integrated criminal justice system, the mental health center case management records, and the county's ambulance transport logs into a single dataset. We formulate our task as a binary classification problem, predicting whether an individual will be booked into jail in the next 12 months. The risk scores generated by the models are then used to prioritize individuals for proactive outreach by mental health professionals. Based on Johnson County's available resources and bandwidth, they can initially intervene on 200 individuals in the next year, leading us to use precision in the top 200 as our primary evaluation.

5 DATA SOURCES

The primary data sources are at an individual level and come from three county departments: Criminal Justice, Emergency Medical Services, and Mental Health.

5.1 Criminal Justice

The Johnson County jail and court system has an integrated justice management system that spans interactions from booking through probation. Data from some services, like jail bookings, span back to 1993, but the dataset is most reliable and complete from 2010 onward. A total of 100 646 jail bookings are recorded in this dataset, with 50 841 unique individuals since January 1, 2010.

Data in the integrated justice management system includes dates of entry and exit, bail amounts and outcome, charges leveled, trial disposition, and probation. Notably absent in this dataset are contacts with local law enforcement that do not result in a jail booking. Individuals with repeated interactions in this dataset are accurately linked upon jail entry by matching fingerprints.

5.2 Emergency Medical Services

County ambulance service is coordinated through a single entity, MED-ACT. The data provided by MED-ACT includes all ambulance dispatches, patient contacts by ambulance paramedics, and transports since March 2010. This dataset includes 200 479 entries for 114 574 unique individuals, of which 80.6% were transported and 84.5% were treated.

Data from Emergency Medical Services includes triage code, patient disposition, and transported destination. Notably missing are calls for service that neither initiated nor required an ambulance dispatch since those are served by emergency medical technicians stationed within municipal fire departments.

Paramedics attempted to link patients in the dataset at the time of service, using name and date of birth. Successful matches were verified by the patient's medical history. Some repeated interactions

were logged without being linked to the same individual; in these cases they were linked in a subsequent analysis. See Section 5.4 for more details.

5.3 Mental Health

The county's public mental health offices provided electronic case files dating from 2010. Prior to that the office used paper records; a small number of those records dating to 1970 have been manually entered into the computer system. Over 1.0 million services have been recorded for 19 751 individuals since January 1, 2010. About 20% of these services are phone calls, with the remainder divided into smaller categories of therapies, care, and transportation.

Of particular interest within this dataset are the diagnoses, programs and dates of services, and discharge reasons. Missing are data from the numerous private behavioral health centers and therapists within the county.

5.4 Combining Datasets

All three datasets were first deidentified by hashing names and social security numbers in a consistent manner prior to analysis. This was done to protect the privacy of individuals in the data while still allowing the linking to take place. Rows from each dataset were linked to the same individual through a combination of probabilistic matching and record linkage. The open source 'dedupe' package [2] was configured to treat the hashes and census tract locations as exact strings, with fuzzy string matching on date of birth, and categorical matching on other demographics such as race and gender. Over 100 conservatively labeled examples provided the training set. Following the probabilistic matching, an exact record linkage step was performed to merge identities that had been identified as the same through the fingerprinting process in the criminal justice dataset.

The process resulted in 127 000 individuals, approximately 10% (12 280) of whom had entries in two or more datasets. Figure 1 demonstrates the resulting sizes and overlaps between the datasets relative to the total county population.

The deidentification process was conservative, making it difficult to find matches across typos and misspellings. We are actively working to deploy a more robust matching algorithm within the protected environment on Johnson County's servers to work directly with the original data.

6 METHODS

The goal of the Early Interventions System (EIS) for Johnson County is to identify individuals who are at risk of going to jail but whose needs would be better served by being redirected to effective mental health services. We formulate the problem as a binary classification problem where the class of interest is whether an individual will enter jail within the next year. Johnson County has resources to reach out to and intervene with 200 high-risk individuals every year. In consideration of this limitation, our system generates a list of 200 individuals ranked by risk scores, which indicate the risk of a person entering jail within the next year. We developed 252 unique features and fit a variety of classification models, using temporal validation [3] to search for a model that performed well and consistently over a 5-year evaluation window.

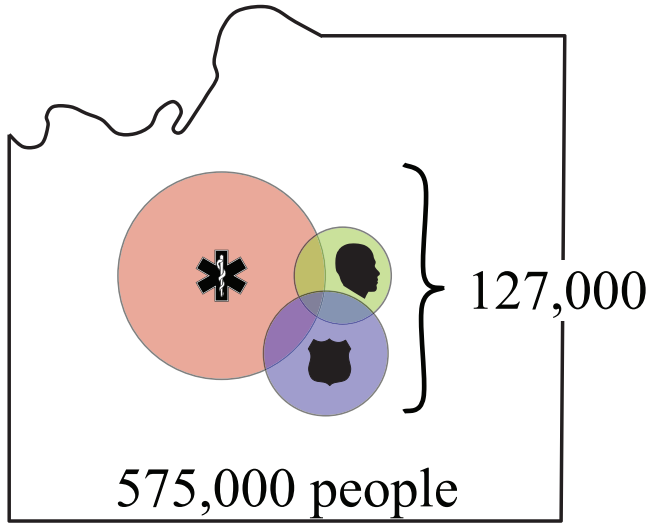


Figure 1: A scale Venn diagram demonstrating the number of individuals in each dataset since January 1, 2010, relative to the total number of residents. The enclosed area represents the entire county, with each overlapping circle representing a public county service.

6.1 Feature Generation

Features were generated based on consultation with various government agencies in Johnson County and our own experience working on data science projects. The data from the three services are similarly structured, capturing detailed information about individual-level demographics and interactions with EMS, criminal justice institutions (jails, courts, and probation), and mental health services. We created three classes of features from these datasets: stable demographics; statistics about the frequency, duration, and number of interactions individuals had with each of the systems; and contextual information about those interactions. Table 1 gives examples of each type of feature.

6.2 Model Fitting

Models were fit using the scikit-learn package in Python [10] and were retrained every year during our evaluation window. Table 2 shows the model and hyperparameter space over which we searched. Our goals with this search was to cast a wide net for a well performing model and to have a number of simple models (shallow decision tree classifiers) to simulate heuristics for baseline comparisons. For the temporal validation process, we trained each model-hyperparameter combinations from Table 1 on data from five time periods. Models were first built using features generated on data as of the beginning of 2010 and labels based on jail bookings in 2010. Then, features from data up to the beginning of 2011 and labels from 2011 were used to retrain the same set of model types. This rolling window process continued through three more years, ending with models trained on features as of the beginning of 2014 and labels generated from 2014 outcomes. Each model was used to generate predictions for outcomes from the following

Table 1: Example Features

Feature Classes and Examples	
Demographics	
	Gender
	Race
	Age
Interaction Statistics	
	Number of bookings ever and in the last year, last month, and last week
	Number of EMS calls ever and in the last year, last month, and last week ever and in the last year, last month, and last week
	Total number of enrollments in any mental health program ever and in the last year, last month, and last week
	Mean and standard deviation of time between interactions with any public service
Interaction Context	
	Age at most recent interaction with any public service
	Age at first interaction with any public service
	Mean amount of bonds paid ever and in the last year, last month, and last week
	Counts of types of bonds (e.g., personal recognizance, cash, surety) ever and in the last year, last month, and last week
	Number time times arrested by specific agencies
	Number of charges of different types (criminal, domestic violence, juvenile, etc.) ever and in the last year, last month, and last week
	Primary impression of paramedic
	Number of unique residential cities given to EMS
	Ever transported to specific hospitals
	Number of EMS calls with specific triage codes
	Number of EMS calls with different outcomes (e.g., refused care, transported)

year (i.e., the 2010 model was used to predict bookings in 2011 and so on).

6.3 Model Selection and Evaluation

Methodology

Based on Johnson County’s capacity for intervention, we analyzed whether the 200 highest scored individuals were booked into jail in the next year. We used *precision at 200* for years 2011 through 2014 to select a final model, and kept the data from 2015 as our final holdout set to evaluate the model selected through this process. Because we are most interested in precision at 200 people, we used the level and stability of this metric over time to select a model that had fairly high and consistent performance in each of the prediction years. Once this model was selected, we examined the performance of this model for predictions in 2015, described below.

For baseline comparisons, we sought to simulate how an expert might make decisions based on simple heuristics—by identifying people at risk through one or two features. We selected simple decision trees of depth 1 and 2 (see Table 1) using the same

Table 2: Grid Search Parameters for Model Selection

Models and Hyperparameters	
Logistic Regression	C: 0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10 Penalty: L1, L2
Random Forest Classifier	Number of Estimators: 1, 10, 100, 1000, 10000 Max Depth: 1, 5, 10, 20, 50, 100 Max Features: Square root, log2 Minimum Samples at Split: 2, 5, 10
K Nearest Neighbors Classifier	N Neighbors: 1, 5, 10, 25, 50, 100 Weights: uniform, distance Algorithm: auto, ball tree, kd tree
Decision Tree Classifier	Criterion: gini, entropy Max Depth: 1, 2 Minimum Samples at Split: 2, 5, 10 Minimum Samples per Leaf: 1, 200
AdaBoost Classifier	Algorithm: SAMME, SAMME.R Number of Estimators: 1, 10, 100, 1000, 10000
SGD Classifier	Loss: hinge, log, perceptron Penalty: L2 L1, Elasticnet
Extra Trees Classifier	Number of Estimators: 1, 10, 100, 1000, 10000 Criterion: gini, entropy Max Depth: 1, 5, 10, 20, 50, 100 Max Features: sqrt, log2 Min Samples Split: 2, 5, 10

model evaluation procedure. In general, we believe it’s important to understand how well simple rule-based systems can perform before deciding to implement a complex machine learning system that is expensive and more difficult to build and maintain for government agencies.

7 RESULTS

The final model selected based on data from 2011 to 2014 was a Random Forest classifier with 100 trees, a minimum of 5 samples at split, square root maximum features used at each split, and a maximum depth of 20. Of the 200 individuals the model identified as highest risk in our holdout year (2015), 102 were actually booked into jail—a precision of 51%.

7.1 Accuracy

Our simulated heuristic models relied on seemingly sensible thresholds but performed consistently worse than the random forest. In 2015, the 1-depth tree (decision stump) classified individuals who had one jail booking in the last year as at risk and had a precision

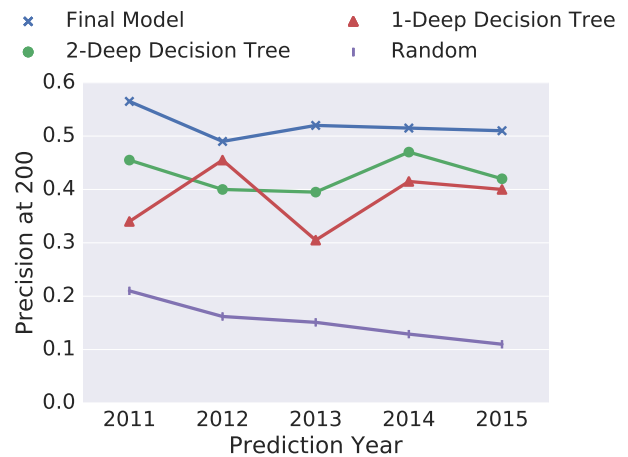


Figure 2: Precision for the top 200 individuals at risk of re-entering jail in each prediction year from our final model, two models simulating simple heuristic decision-making (decision trees), and the expected value of random selection.

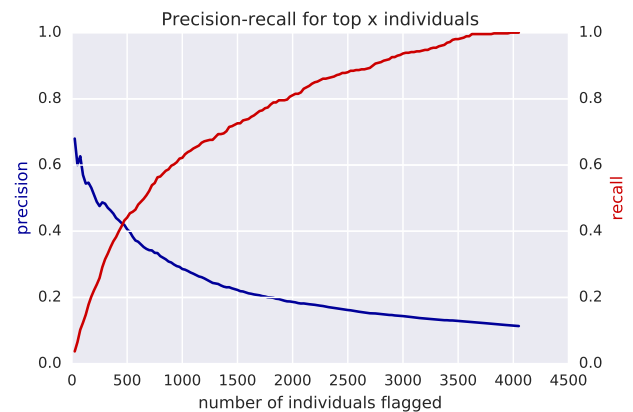


Figure 3: Precision and recall of the final model in 2015 when the threshold for the positive (jail booking) label is set at the xth highest risk score.

of 40% for 200 individuals selected from this group. In the same year, the 2-depth tree performed slightly better (precision 42%) by including whether the average time between a person’s interactions with any of the three public services exceeds 611 days. Both of these models showed higher variance over the five evaluation years than the random forest model, as shown in Figure 2.

Overall, our model outperforms both random selection and simple heuristics at identifying people with mental illness at risk of re-entering jail. Figure 3 shows that, although the precision of our model decreases as more individuals are labeled at risk, it remains much higher than chance even if the capacity for intervention is doubled to 400 individuals or more.

While our model is 20%–25% more precise than the two heuristic models for 2015, its relative performance is even better when

looking at length of jail stays. As shown in Figure 4, the higher an individual's risk (as estimated by our model), the more days in jail he/she tends to have. The individuals we correctly predicted would re-enter jail in 2015 collectively had 183 jail bookings (nearly 2 each on average) and stayed in jail for a total of 6 762 days—about 18 years in total, or a little over two months per person. That is 37% longer than the jail stays for the two-deep decision tree's top 200 and 47% longer than the jail stays for the one-deep decision tree's. This confirms that our model more accurately flags the highest-risk individuals, thereby giving officials the opportunity to target their limited resources to those who need it most.

Figure 5 shows cumulative days spent in jail by risk score and further illustrates that the majority of these days are accrued by a relatively small number of individuals. This gives us additional confidence that the model accurately identifies individuals who are at risk of spending significant amounts of time in jail as opposed to short stays. The Figure also helps characterize the population that Johnson County might intervene on if its resource constraints were to change. While the cumulative number of jail days potentially affected would decrease if the County decided to intervene on fewer than 200, it could also potentially affect roughly 3 000 more jail days by intervening on the 100 next riskiest individuals.

7.2 Exploring the (Predicted) High Risk Individuals

The top 200 individuals as identified by the model differed from the rest of the population across several key metrics and demographics. One of the largest disparities was in the average number of days between interactions with any of the three public services; the top 200 individuals averaged just over a year (360 days) between interactions whereas the rest of the population averaged nearly two and a half years (908 days). While the high-risk population had nearly twice as many services (112) recorded by the mental health centers as the others (63), they also tended to have a longer time since their last contact with the mental health center (median of 1 036 days compared to 790).

In the criminal justice system, there was a large disparity between the average bail amounts over the previous year between the two populations. The higher risk group averaged \$6 270 as compared to \$3 950 for the others.

The top 200 also differ from the rest demographically. Although the majority of criminal justice contacts involve men, the proportion of within the top 200 was higher (72% compared to 61%). About 16% of both groups are African American, whereas white people are more highly represented in the top 200 (79% versus 55%). The top 200 were substantially younger than the rest of the population, averaging over 7 years younger (26.9 years of age as compared to 34.2).

8 IMPLEMENTATION PLAN AND NEXT STEPS

Johnson County officials are planning to use this model to conduct proactive mental health outreach. In parallel, the system is being extended to larger populations and more outcomes to address additional needs in the community. As a first step, Johnson County is evaluating the lists of individuals our model deemed highest risk.

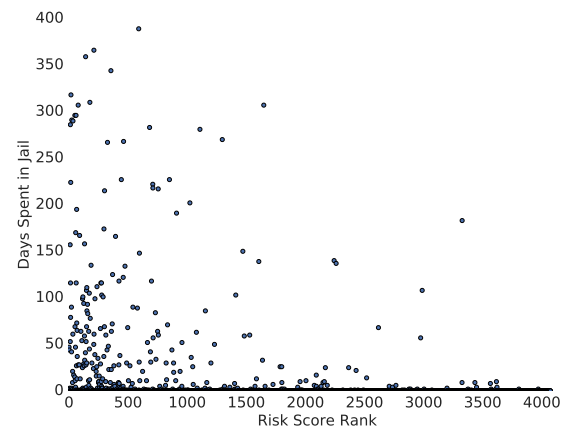


Figure 4: Scatter plot of days spent in jail as a result of bookings in 2015 and ranking in risk scores from our final model (low numbers indicate greater risk). Most individuals spent less than one day in jail (either were never booked or were booked and released on the same day); however, those who experienced longer incarcerations generally ranked higher on risk scores.

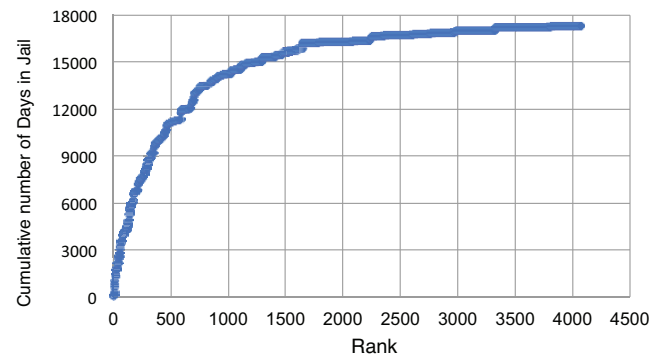


Figure 5: Cumulative days spent in jail as a result of bookings in 2015. Lower ranks indicate greater risk and higher priority.

This process involves discussions with stakeholders, including Mental Health Center employees, about the validity of entity matching, the usefulness of the lists, and the potential for interventions.

The interventions stemming from this work include outreach by mental health caseworkers to individuals identified as at risk by our model and proactive follow-ups by mental health caseworkers with individuals who have been out of contact for a substantial amount of time (e.g., over one year) in order to maintain accurate contact information and reassess current needs. Given that the median time since last contact for the at-risk population was over eight months longer than the rest of the population, this suggests a positive impact from interventions designed to reconnect individuals at risk to mental health services.

Implementation of any model-based interventions will involve training program administrators on the interpretation and appropriate use of risk scores as well as a monitoring and evaluation plan to assess program impact.

Future work includes modeling the length of jail time directly to provide another set of predictions to criminal justice agencies. We are also continuing to expand this work by integrating new sources of data (e.g., substance abuse, homelessness, and police data) and working with additional jurisdictions throughout the United States. This allows us to continue to improve our models and to understand the generality of this method and our findings across not only additional locations but also additional outcomes we want to prevent. We plan to build additional models to predict other risky, complex patterns of social service interactions such as :

- when individuals with mental illness are at risk of requiring emergency medical services,
- an individual's risk of jail booking or need for mental health services when they call emergency medical services, or
- when someone may be at risk of dropping out of mental health services prematurely.

The prioritized rankings provided by these models will enable outreach programs wherein individuals' underlying vulnerabilities can be identified and mitigated by directing them to appropriate services like housing assistance, mental health triage, or substance abuse counseling.

Another important area of future work is to incorporate notions of bias and discrimination in the model predictions. This is especially important in criminal justice where a lot of historical data comes from a biased process and using that data to build machine learning models may lead to reinforcing that bias. Since our work is intended to provide additional support to individuals who may be at risk of jail interactions, we want to make sure we are not missing individuals due to biases in our data. It's also important, for resource allocation purposes, to have a low false positive rate. We are currently exploring methods to *audit* predictions provided by machine learning models to understand and highlight potential bias as well as provide tools to policymakers to correct for that bias.

9 CONCLUSION

Using historical data from ambulance transports, behavioral health services, and the criminal justice system, we have built a machine learning system that predicts when individuals with histories of mental illness and incarceration are at risk of re-incarceration. Preliminary evaluations of this model using temporal cross-validation demonstrate significant advantages over random selection or simpler heuristics. Over half of the 200 individuals identified as being most at risk of re-incarceration at the beginning of 2015 spent time in jail later that year, combining for nearly 18 years of jail time. This represents a significant avenue for Johnson County to implement an intervention program where human insights into the ongoing struggles can provide substantive help in getting people connected with comprehensive care, counseling, or treatment.

ACKNOWLEDGMENTS

We would like to thank the Data Science for Social Good program and the Center for Data Science and Public Policy at the University

of Chicago for its generous support of this work and the Obama White House Office of Science and Technology Policy for their help and support. We thank the County Manager's office of Johnson County as well as the many individuals in Johnson County that made this work possible, including (but not limited to) Tim Mulcahy and Songying Yang from Justice Information Management Systems, Dr. Ryan Jacobsen, Medical Director, Kimberly Rowland and Megan Younger, Co-responders, and Tim DeWeese, Director of the Mental Health Center.

REFERENCES

- [1] Karen M Abram, Linda A Teplin, and Gary M McClelland. 2003. Comorbidity of severe psychiatric disorders and substance use disorders among women in jail. *American Journal of Psychiatry* 160, 5 (2003), 1007–1010.
- [2] Forest Gregg and Derek Eder. 2016. Dedupe. <https://github.com/datamade/dedupe>. (2016).
- [3] Rob J Hyndman and George Athanasopoulos. 2014. *Forecasting: principles and practice*. OTexts.
- [4] Doris J James and Lauren E Glaze. 2006. Mental Health Problems of Prison and Jail Inmates. (2006).
- [5] Travy Johnson, Debra Rinehart, Josh Durfee, Holly Batal, Joshua Blum, Carlos Oronce, Paul Melinkovich, and Patricia Gabow. 2015. For many Patients Who Use Large Amounts of Health Care Services, The Need is Intense Yet Temporary. *Health Affairs* 34, 8 (2015), 1312–1319.
- [6] Karen E Joynt, Atul A Gawande, John Orav, and Ashish K Jha. 2013. Contribution of Preventable Acute Care Spending to Total Spending for High-Cost Medicare Patients. *Journal of the American Medical Association* 309 (2013), 2572–2578. Issue 24.
- [7] H Richard Lamb and Linda E Weinberger. 1998. Persons with severe mental illness in jails and prisons: A review. *Psychiatric services* (1998).
- [8] Mark R Munetz and Patricia Griffin. 2006. Use of the sequential intercept model as an approach to decriminalization of people With serious mental illness. *Psychiatric services* 57 (2006), 544–549. Issue 4.
- [9] Council of State Governments/Eastern Regional Conference and United States of America. 2002. Criminal Justice/Mental Health Consensus Project. (2002).
- [10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [11] E Fuller Torrey, Aaron D Kennard, Don Eslinger, Richard Lamb, and James Pavle. 2010. More Mentally Ill Persons Are in Jails and Prisons Than Hospitals: A Survey of the States. (2010).