Implementing a Predictive Risk Model to Prioritize Homeless Services: The Allegheny Homelessness Tool

CENTRE FOR SOCIAL DATA ANALYTICS

Chamari I. Kithulgoda*, Emily Kulick†, Abigail Horn‡, Erin Dalton‡, Sophie Joyce* and Rhema Vaithianathan*

*Centre for Social Data Analytics, Auckland University of Technology, Auckland 1010, New Zealand *Email: {chamari.kithulgoda, sophie.joyce, rhema.vaithianathan}@aut.ac.nz*

†Analytics for Change, Chicago, Illinois 60647, USA *Email: emily@analyticschange.com*

‡Allegheny County Department of Human Services, Pittsburgh, PA 15222, USA *Email:{Abigail.Horn, Erin.Dalton}@alleghenycounty.us*

Background

With the recent rise in homelessness across the US[1], demand for homelessness programs outstrips capacity. This means eligible people must be prioritised. Most jurisdictions use some form of prioritisation mechanism - principally actuarial tools.

Located in Pittsburgh, PA, the Allegheny County Department of Human Services (DHS) is unique in having invested considerable resources over the last 20 years in an integrated data warehouse. This poster reports on work with the DHS to use that data to replace the homeless services actuarial tool with a predictive risk model (PRM).

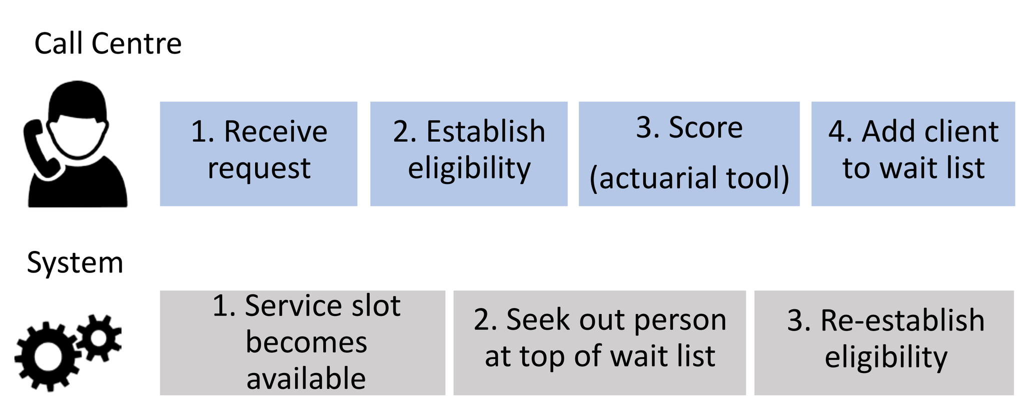
Obtaining social licence

The DHS is committed to transparent and ethical implementation of machine learning tools. The inputs into gaining social licence are illustrated below.

How are homeless services currently allocated?

Allegheny DHS receives over 10,000 requests for homeless services annually. It runs a continuum of care model offering services including homeless support, homeless prevention, street outreach, emergency shelter, bridge housing, permanent supportive housing and rapid rehousing.

At any time there are around 550 households waiting for service slots - but only 50 slots come free each month. This means systematic prioritisation is required.



How does the actuarial tool work?

The actuarial score is calculated based on the client's response to a series of questions asked at the call centre. Answers are weighted by a predefined matrix of weights.

Why did DHS decide to explore the implementation of a PRM tool?

The actuarial tool requires clients to reconstruct their history which can be difficult and traumatic. The DHS already has much of the data, and could therefore eliminate or reduce the need for self reported data.

The DHS also wanted to test whether a PRM tool could better identify clients at risk of harmful outcomes, compared with the actuarial tool.

Building the PRM

Objective of PRM

Prioritise supportive housing for people who are at highest risk of harms associated with homelessness.

Data

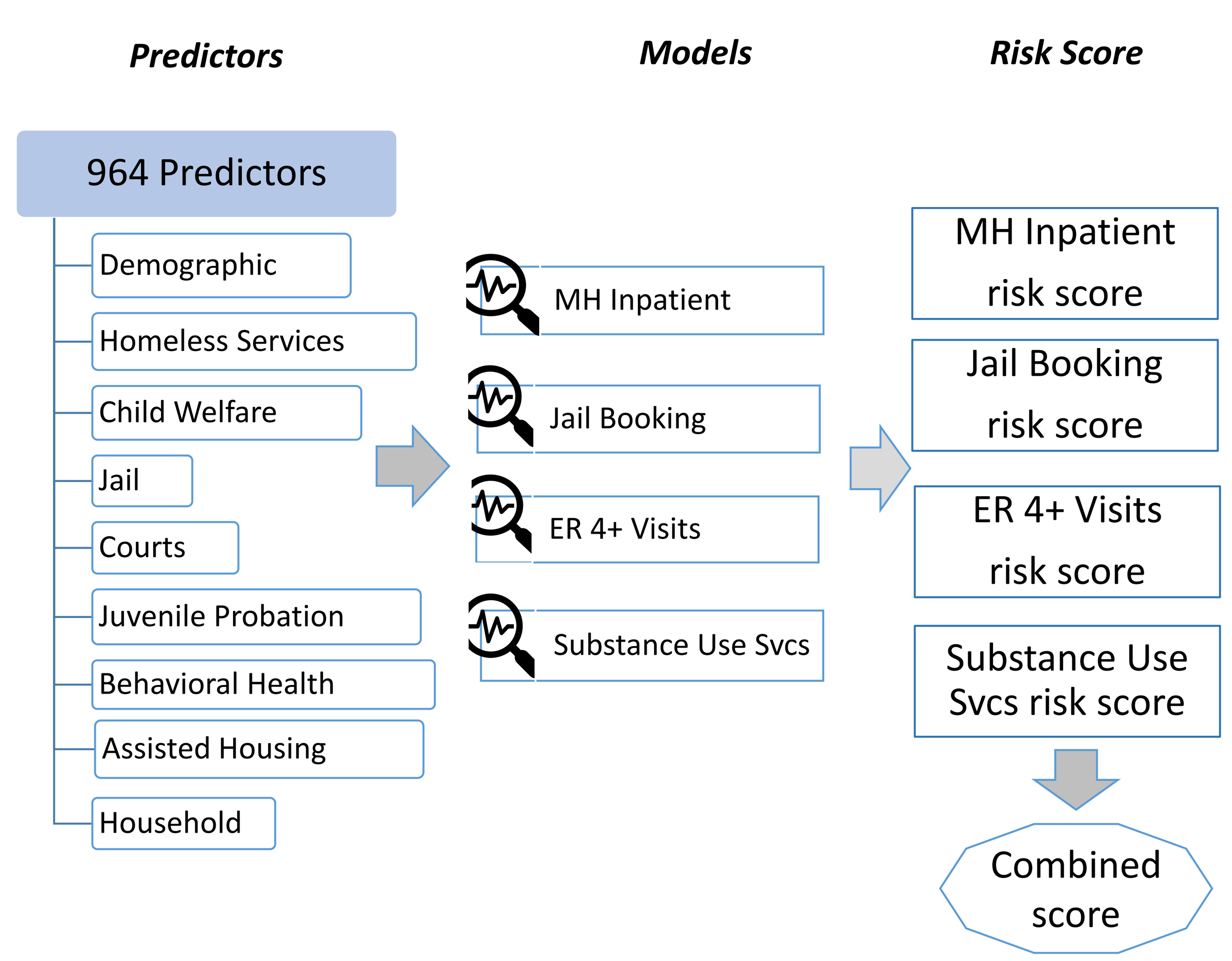
Research data included 5,550 observations from assessments conducted by Allegheny DHS between January 2016 and March 2017.

We trained models on four target outcomes (described below), using 964 predictors from nine domains.

Model	Description of Target Outcome	Prevalence
MH Inpatient	At least one inpatient mental health service in the 12 months following the call	15.98%
Jail Booking	At least one Allegheny County Jail booking in the 12 months following the call	17.11%
ER 4+ Visits	Four or more ER visits in the 12 months following the call	20.52%
Substance Use Svcs	At least one substance use services contact in the 12 months following the call	29.14%

Modeling

LASSO regularized Logistic Regression was the machine learning algorithm of choice. Each model was instantiated through the R package glmnet[2]. Predicted risk probabilities were grouped into 20 equal–sized bins defined by quantiles.



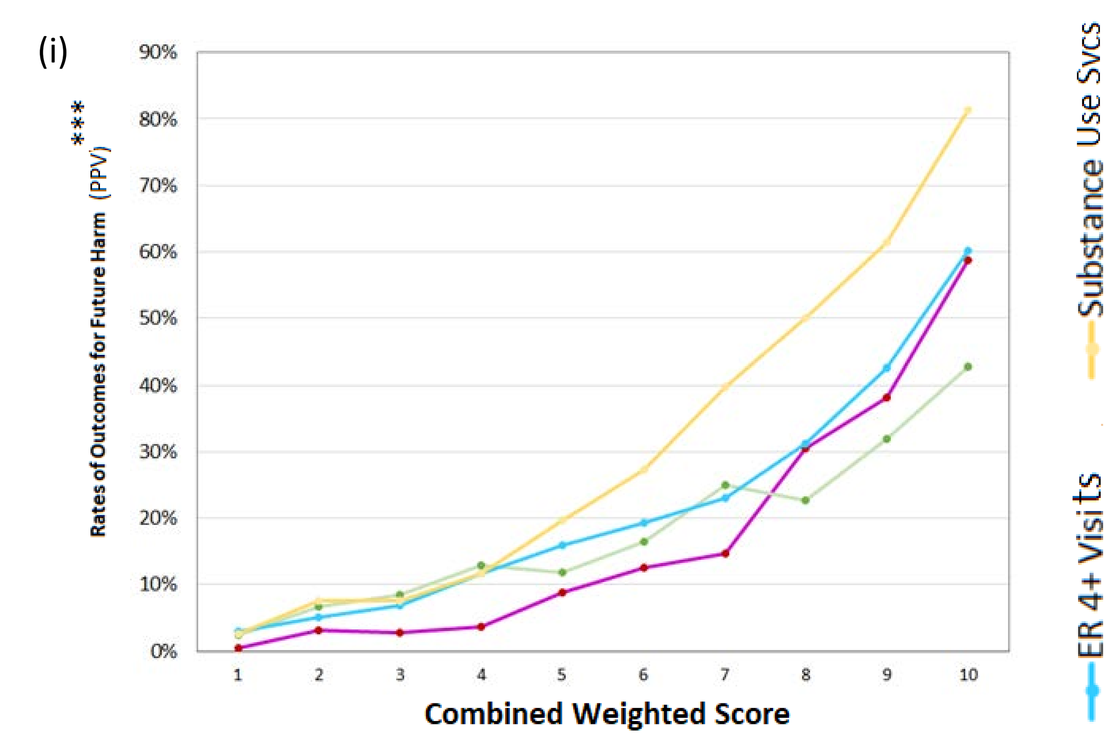
Performance Evaluation

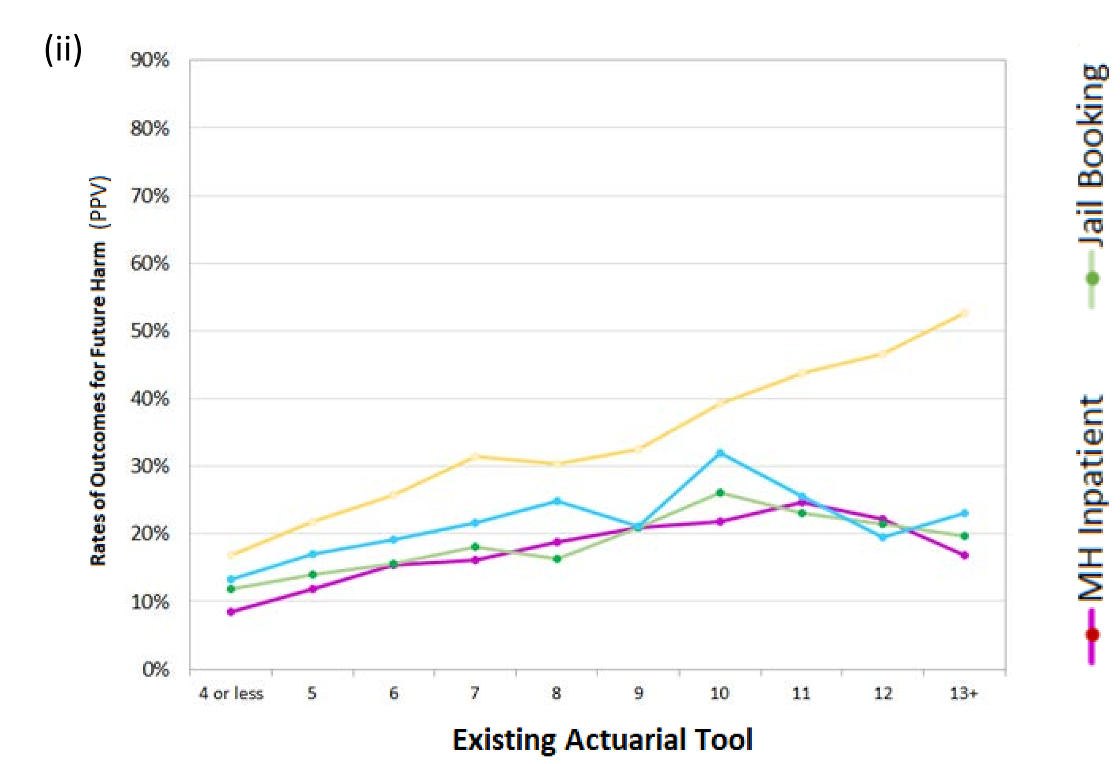
We reported results for two different approaches: individual score and combined weighted score.*

The table below reports the predictive performance of each model against trained and non-trained outcomes.

Approach		AUROC** across all trained and non-trained outcomes			
		MH Inpatient	Jail Booking	ER 4+ Visits	Substance Use Svcs
Individual Score	MH Inpatient	**83.58%**	66.85%	76.58%	77.46%
	Jail Booking	71.08%	**74.19%**	61.60%	71.77%
	ER 4+ Visits	79.28%	62.40%	**81.29%**	73.60%
	Substance Use Svcs	78.16%	68.37%	71.28%	**85.97%**
Combined Weighted Score		83.03%	72.46%	77.35%	82.43%

The following graphs illustrate the prevalence of target outcomes against (i) the combined weighted score and (ii) the existing actuarial tool.





** Risk scores for MH Inpatient, Jail, ER 4+ visits, and Substance Use Svcs were weighted so as the highest score of 20 was replaced by 5, 18-19 by 4, 15-17 by 3, 10-14 by 2, 5-9 by 1 and scores less than 5 by 0. Then the total was recalibrated for 10 equal bins.*

*** Area Under the Receiver Operating Characteristics curve*

****Positive Predictive Values*

Validation & Discussion

External Validation

The performance of the models was also tested against two additional non-trained outcomes: mortality and chronic homelessness.

Non-trained Outcome	Description	Prevalence
Mortality	Recorded death in the 12 months following the call	1.50%
Chronic Homelessness	An interaction with emergency shelter or street outreach in four of the 12 months following the call	10.47%

We compared the risk of those who were scored in the top 10% with the rest to measure relative risk.

Scoring Technique	Relative Risk of Mortality (Top 10%)			Relative Risk of Chronic Homelessness (Top 10%)		
	All	Black Only	Non-Black Only	All	Black Only	Non-Black Only
MH Inpatient	3.25	4.66	2.10	1.29	0.99	1.36
Jail Booking	1.83	1.41	2.00	1.06	0.75	1.27
ER 4+ Visits	2.86	2.53	2.58	1.48	0.97	1.71
Substance Use Svcs	1.67	1.36	1.37	1.13	0.67	1.19
Combined Weighted Score	**3.59**	**3.79**	**2.85**	1.36	0.71	1.68

Discussion

Overall, the PRMs performed well, even on harms they were not trained for. By contrast, the scores of the actuarial tool were only weakly related to the future harms that we studied.

Because the models are trained on specific harms that are identifiable in the administrative data, we also tested how well the models were able to predict the more objective harm of death. Overall, those in the top 10% of the combined score were ~3.6 times more likely to die - and this rate was fairly similar across Black and Non-Black populations.

Implementation of the PRM tool will provide better targeting of restricted homelessness services to those who are most likely to be subject to future harms. It will also reduce the amount of time that DHS staff spend assessing clients.

The PRM tool is expected to be deployed by early 2020.

Acknowledgements

The authors wish to express their thanks for the financial support of Allegheny County DHS. We would also like to show our gratitude to Andy Halfhill from Allegheny DHS for sharing his knowledge and insights that assisted this work. We thank Matthew Walsh and Oleksandr Fialko from Centre for Social Data Analytics, Auckland University of Technology for their contribution on the earlier versions of this project.

References

1. The U.S. Department of Housing and Urban Development-Office Of Community Planning and Development.(2019). The 2018 Annual Homeless Assessment Report (AHAR) to congress. [Online]. Available: https://files.hudexchange.info/resources/documents/2018-AHAR-Part-1.pdf
2. J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," Journal of Statistical Software, vol. 33, no. 1, pp. 1–22, 2010. [Online]. Available: http://www.jstatsoft.org/v33/i01/