

NIJ Recidivism Forecasting Challenge Webinar

May 6, 2021



Presenters

Joel Hunt
Senior Computer Scientist

Angela Moore
Senior Science Advisor

Marie Garcia
Senior Social Science Analyst

Eric Martin
Social Science Analyst

D. Michael Applegarth
Research Assistant

National Institute of Justice (NIJ)

- Research, development, and evaluation agency of the U.S. Department of Justice
- Dedicated to improving knowledge and understanding of crime and justice issues through science
- Invests in scientific research across disciplines to serve the needs of the criminal justice community



Data Science Research at NIJ

- Principal focus on analysis and visualization of crime data
- Facilitated implementation of computer-based crime mapping and analysis in American policing



Data Science Research at NIJ

- 1986: First grant award to University of Illinois at Chicago to explore mapping crime in the context of community policing
- 1990: Initiated the Drug Market Analysis Program
- 1997: Funded the development of CrimeStat
- 1997: Established the Crime Mapping Research Center (later, the Mapping and Analysis for Public Safety Program)



Data Science Research at NIJ

- 2009: Began exploring the potential of crime prediction/forecasting
- 2016: Issued the Real-Time Crime Forecasting Challenge
- 2021: Issued the Recidivism Forecast Challenge



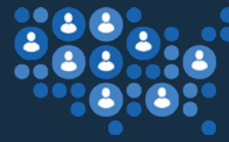
Goals of the Challenge

- Enhance recidivism forecasting using person- and place-based factors
- Understand how supplemental data can be integrated with official records
- Increase accuracy of risk predictions for all individuals under supervision



Why do We Need the Challenge?

- It is critical for practitioners to know who is most likely to recidivate
 - Allows them to triage within risk categories in their caseload in order to provide tailored services to those who need them the most
- Congruence of predicted risk to actual risk increases likelihood of success
 - High-intensity supervision for low risk offenders may increase recidivism



Why do We Need the Challenge?

- Meet the needs of all supervisees
 - Conventional risk assessment tools predict risk better for white men
 - The Challenge structures prizes to encourage robust predictions that are accurate across racial groups
 - Community supervision needs to understand and meet the unique needs of female probationers and parolees
 - The Challenge encourages submission of different models for male and female parolees



Challenge Dataset

- Uses data from the Georgia Department of Community Supervision and the George Crime Information Center
- Contains data for over 25,000 individuals released from prison to parole from January 1, 2013 through December 31, 2015
- Data will be released in 3 waves (initial data release, year 2 update, & year 3 update)
- 70% of the dataset will be released as training data (i.e., all dependent variables will be included) and 30% will be released as test data



Challenge Codebook

- Challenge Dataset contains variables in the following categories:
 - Supervision Case Information
 - Prison Case Information
 - Prior Georgia Criminal History
 - Prior Georgia Community Supervision History
 - Georgia Board of Pardons and Paroles Conditions of Supervision
 - Supervision Activities
 - Recidivism Measures



Utilizing Additional Data

- Contestants may use any other data sets or services to enhance the effectiveness or efficiency of their forecasts.
 - Get creative!
- Potential supplemental data resources are provided in Section IV of the Challenge document, including:
 - National Archive of Criminal Justice Data (NACJD)
 - Census Bureau's American Community Survey (ASC) Public Use Microdata (PUMS)



Entries

- Contestants **must** submit entries under the appropriate contestant category:
 - **Students** —Enrolled as a full-time student in high school or as a full-time, degree seeking student in an undergraduate program.
 - **Small Teams / Businesses** —Team comprised of 1–10 individuals, or a small business with less than 11 employees.
 - **Large Businesses** —A business with 11 or more employees.
- Entries **must** include a team roster.



Entries

- Entries should **not** be submitted in a zipped folder
- The forecast files **must** be named with the following **naming** convention:
“TEAMNAME_1YearForecast”
“TEAMNAME_2YearForecast”
“TEAMNAME_3YearForecast”
- Variable fields **must** be named “ID” & “Probability”
- The same naming convention should be followed for the team roster:
“TEAMNAME_1YearRoster”
“TEAMNAME_2YearRoster”
“TEAMNAME_3YearRoster”

Requirement for Entries	
Requirement	Description of Requirement
File formats (must be machine readable)	.xls .xlsx .csv .txt .odf
Required variables	Maintain Original ID field provided Probability Field, between 0 and 1 with up to 4 decimal places



Entries

- Registration and entry are free
- Registration and submission instructions are on the Challenge website: nij.ojp.gov/funding/recidivism-forecasting-challenge



Eligibility

- The Challenge is open to residents of the United States and its territories who are 13 years and older at the time of entry.
 - Entries by contestants under the age of 18 must include the co-signature of the contestant's parent or legal guardian.
 - Companies must have an office with a U.S. business license.
- Employees of NIJ and individuals or entities listed on the Federal Excluded Parties List are ineligible to participate.
 - Employees of other federal agencies should consult with their Ethics Officer concerning their eligibility.



Judging Criteria

- NIJ will use the Brier Score (BS) metric to measure accuracy of the forecasts:

$$\text{Brier Score} = \frac{1}{n} \sum_{t=1}^n (f_t - A_t)^2$$

- Since the Brier score is a measure of error, applicants should look to minimize this metric (lowest score wins).



Judging Criteria

	\hat{Y} – Fail	\hat{Y} – Succeed	Conditional Procedure Error
Y – Fail	A True Positive	B False Negative	$B/(A+B)$ False Negative Rate
Y – Succeed	C False Positive	D True Negative	$C/(C+D)$ False Positive Rate
Conditional User Error	$C/(A+C)$ Failure Prediction Error	$B/(D+B)$ Success Prediction Error	$(C+B)/(A+B+C+D)$ Overall Procedure Error

To generate a measure of accurate and fair, NIJ will use a confusion matrix using a ≥ 0.5 cut point.

Here Y is the actual outcome and \hat{Y} is the forecasted outcome. A “fail” is considered a positive outcome (the variable of interest occurred, i.e., the individual did recidivate); and, a “succeed” is considered a negative outcome (the variable of interest did not occur, i.e., the individual did not recidivate).

Based on this matrix, a fairness penalty (FP) function is calculated using the racial difference in false positive rates

$$FP = 1 - |FP_{Black} - FP_{White}|$$



Judging Criteria

- Since the Brier Score is a measure of error on predictions bound between 0 and 1, the error should also be bound between 0 and 1. As such we consider one minus the Brier Score a metric of correctness. The final index NIJ will use to calculate which algorithms are the most accurate while accounting for bias is:

$$\textit{Fair and Accurate} = (1 - BS)(FP)$$

- Since this is now a measure of fairness and accuracy, applicants should look to maximize this metric (highest score wins).



Prizes

- Up to 114 prizes will be awarded
 - Each contestant/team may win up to 15 prizes
- The most accurate forecasts for males, females, and a combined score, at each Challenge period (i.e., year 1, year 2, year 3) will be awarded prizes
- Additionally, the forecasts with the highest “fair and accurate” score for the male and female categories, for each Challenge period, will qualify for an additional prize



Prize Structure

Large Team									
	Male Parolees			Female Parolees			Average		
	1 Year Forecast	2 Year Forecast	3 Year Forecast	1 Year Forecast	2 Year Forecast	3 Year Forecast	1 Year Forecast	2 Year Forecast	3 Year Forecast
1 st Place	\$25,000	\$25,000	\$25,000	\$25,000	\$25,000	\$25,000	\$25,000	\$25,000	\$25,000
2 nd Place	\$12,000	\$12,000	\$12,000	\$12,000	\$12,000	\$12,000	\$12,000	\$12,000	\$12,000
3 rd Place	\$4,000	\$4,000	\$4,000	\$4,000	\$4,000	\$4,000	\$4,000	\$4,000	\$4,000
4 th Place	\$2,000	\$2,000	\$2,000	\$2,000	\$2,000	\$2,000	\$2,000	\$2,000	\$2,000
5 th Place	\$1,000	\$1,000	\$1,000	\$1,000	\$1,000	\$1,000	\$1,000	\$1,000	\$1,000
Accounting for Racial Bias									
	Male Parolees			Female Parolees			Average		
	1 Year Forecast	2 Year Forecast	3 Year Forecast	1 Year Forecast	2 Year Forecast	3 Year Forecast	-	-	-
1 st Place	\$12,500	\$12,500	\$12,500	\$12,500	\$12,500	\$12,500	-	-	-
2 nd -5 th place in this prize category will be recognized on the Challenge website.									

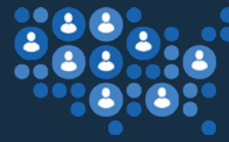


Winners' Paper

- Paper requirements
 - A comprehensive document describing lessons learned, the variables found to be significant predictors of recidivism in the forecast, and the type(s) of models found to outperform others
- Contestants are encouraged to provide
 - Potential measures that could be used in the future, and any additional intellectual property such as: specific techniques, weighting, other sensitive decisions, and if possible uploading their code to an open source platform

Key Dates

- April 30, 2021: Initial release of training data
- April 30, 2021: Initial release of test data
- May 31, 2021: End of submission period 1
- June 1, 2021: Release of updated test data (i.e., year 2)
- June 15, 2021: End of submission period 2
- June 16, 2021: Release of final test data (i.e., year 3)
- June 30, 2021: End of submission period 3
- On or by August 16, 2021: NIJ will announce Challenge winners
- Before September 17, 2021: Winners must submit paper to NIJ



Questions?

- For substantive questions about the Challenge, e-mail NIJ at NIJRecidivismChallenge@usdoj.gov
- For technical questions, contact the OJP Grants Management System Help Desk at 1-888-549-9901 or by email at OJP.ITServiceDesk@usdoj.gov



Q&A

- We will now begin the Q&A