NCJRS
National Criminal Justice Reference Service

Office of Justice Programs
BJA BJS NIJ OJJDP OVC SMART

The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title:	NIJ Recidivism Challenge Report, Team Klus
Author(s):	Team Klus
Document Number:	305040
Date Received:	July 2022
Award Number:	NIJ Recidivism Forecasting Challenge Winning Paper

This resource has not been published by the U.S. Department of Justice. This resource is being made publicly available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

NIJ Recidivism Challenge Report

Team Klus

August 30, 2021

Introduction

Project objective

The objective of this exercise was to develop a model for first-year recidivism that would minimize the potential for racial and gender bias.

Variables

Data source

The data were provided by the National Institute of Justice in conjunction with the Georgia Department of Community Supervision regarding demographics, criminal history, and other characteristics of parolees in the State of Georgia from January 2013 through December 2015. Notably, the data were limited to parolees whose race was identified as either black or white. Parolees of Hispanic, Asian, Native American, or other racial or ethnic groups were not included in the provided data.

External data

For this model, we exclusively relied upon the data provided by the NIJ. While we engaged in some feature engineering (i.e. transformation of variables to achieve stronger or more useful predictors), the data itself was not augmented by any external sources.

Data cleaning & feature engineering

One of the most time-consuming tasks in the development of this model was cleaning the data so that it would be usable for modeling purposes. Re-leveling categorical variables, collapsing categories, and in some cases transforming categorical variables to ordinal variables in order to achieve a more parsimonious model

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

were key components of this task.

We chose to transform variables like prior arrest episodes for felonies and misdemeanors, number of prior arrests for specific categories of crimes, prior number of convictions for specific categories of crimes, and age at release from categorical to ordinal variables to preserve degrees of freedom in our candidate models. The validity of this approach was assessed during exploratory data analysis.

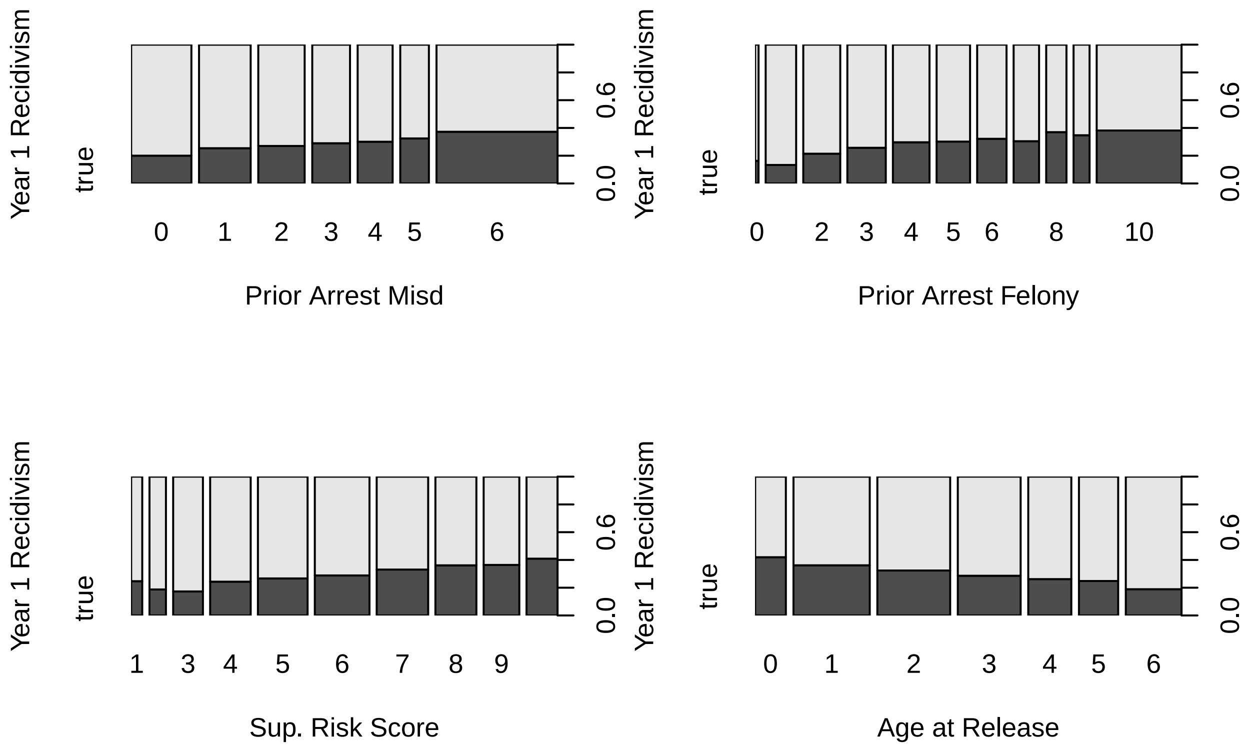
Exploratory data analysis

The most important component of our model development strategy for this challenge was exploratory data analysis. The NIJ data consisted of 49 candidate predictors. To achieve the most strongly predictive, parsimonious model, we needed to reduce the set of candidate predictors by assessing their relationship to the outcome of interest, recidivism, as well as each other.

We initially examined the bivariate summaries of the data to determine the relationship between the candidate predictors and the outcome of interest, recidivism at one year post-release. We viewed the candidate predictors as belonging to one of six broad groups: demographics, general risk factors, criminal history, parole history, drug use, and employment. These variables are summarized in Table 1, and further expounded upon below.

With regard to demographics, females had a lower frequency of recidivism than males in this data (20.6% for females versus 31.1% for males), but they also comprised only 12.3% of the population of parolees in the sample. With regard to race, it appeared from the data that white parolees had a slightly lower frequency of recidivism than black parolees (28.2% versus 31.0%), but that black parolees were overrepresented in this sample. For age, there appeared to be a linear trend whereby older parolees had a lower incidence of recidivism on average. This was an indication that our decision to treat age at release as an ordinal variable was acceptable. There did not appear to be any strong trends with the US Census Bureau PUMA groups for parolee residence, though there was certainly some heterogeneity in recidivism frequency among PUMAs. Education also appeared to be a factor, since parolees with at least some college education exhibited a lower frequency of recidivism than those with a high school education or less (22.3% versus 31.4%). Notably, individuals with at least some college education made up a small proportion of the overall sample at only 17.7%.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.



For general risk factors, those parolees who recidivated in the first year had, on average, a higher risk score assigned by the State of Georgia upon their release from prison (6.58 versus 5.85, scale from 1-10). Supervision level also appeared to be associated with recidivism, although it was not clear how closely associated the risk score and supervision level determinations might be. Finally, 46.9% of those determined to have a gang affiliation recidivated in the first year, compared with 27.7% of those with no known gang affiliation.

For criminal history, inmates with more prior misdemeanor and felony arrests had a higher frequency of recidivism. This trend also held true for misdemeanor and felony convictions. With regard to the type of crime committed, parolees who committed crimes classified as violent sex crimes had the lower frequency of recidivism, at just 10.5%, but these were very rare relative to other classifications, comprising only 3.7% of total prison offenses in this sample. Those imprisoned for property crimes had the highest rate of recidivism, at 35.4%, and this was also the most common type of offense, comprising some 36.9% of the sample. Violent non-sexual offenders had a 26.5% recidivism rate and comprised 24.2% of the sample, drug offenders had a 25.9% recidivism rate and comprised 23.0% of the sample, and finally all other crimes not falling into the previous categories comprised 12.2% of the sample and these offenders had a 31.9% recidivism rate. Prior convictions for a variety of crimes and categories of crimes were also part of the data, and included violent

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

crime, property crimes, drug crimes, parole or probation violations, domestic violence, and gun crimes. These data all followed the same general trend whereby the more instances of prior convictions for each crime, the higher the average frequency of recidivism within one year of release. These variables also appeared to in general follow a linear trend, which supported our decision to treat them as ordinal variables in order to reduce the number of coefficients we would need to estimate in our candidate model.

Table 1: Prison offenses conditional on gender.

	Drug	Other	Property	Violent/Non-Sex	Violent/Sex
F	0.31	0.08	0.44	0.17	0.00
M	0.22	0.13	0.36	0.25	0.04

For variables related to parole history and violations, the data were somewhat counterintuitive in that parolees with recorded infractions, including violations of electronic monitoring, instructions, failure to report, and moving without permission, appeared to have had slightly lower rates of recidivism. Delinquency reports and program attendances appeared to follow a similar trend, where more delinquency reports and greater attendance at required programs were associated with a greater frequency of first-year recidivism. Having many residency changes was also associated with a higher rate of first-year recidivism, but it appeared from the code book that this variable considered residency changes over the entire length of an individual's parole supervision, not just the first year.

For variables related to drug testing, it appeared that in general the distribution for those that recidivated in the first year had longer upper tails for THC and methamphetamine, but not necessarily cocaine. There did not appear to be a strong association between first-year recidivism and the average number of days between drug tests.

Finally, for variables related to employment, parolees who did not recidivate in the first year were associated with a much higher percentage of days employed. Again, this variable appears to capture employment status over the course of the entire length of their parole, and not necessarily just the first year. Parolees who averaged more jobs per year also appeared to be associated with a lower frequency of first-year recidivism, but the exact implication of this variable was not clear. Parolees who were exempt from employment made up a very small proportion of the overall parolee sample, but did appear to have a slightly lower frequency of first-year recidivism. There is likely some other characteristic of this subset of the sample that might explain why they were exempt and why they appeared less likely to recidivate.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 2: Univariate summary statistics stratified by year one recidivism.

	false	true
n	12651	5377
Gender (%)	10891 (86.1)	4920 (91.5)
Race (%)	5536 (43.8)	2179 (40.5)
Age_at_Release (mean (SD))	3.10 (1.89)	2.52 (1.80)
Gang_Affiliated (%)	1477 (13.6)	1304 (26.5)
Supervision_Risk_Score_First (mean (SD))	5.85 (2.38)	6.58 (2.32)
Supervision_Level_First (%)		
High	3237 (27.4)	1666 (33.4)
Specialized	3417 (28.9)	1525 (30.6)
Standard	5177 (43.8)	1794 (36.0)
Education_Level (%)		
<HS	4714 (37.3)	2170 (40.4)
College	2484 (19.6)	714 (13.3)
HSgrad	5453 (43.1)	2493 (46.4)
Dependents (%)		
0	3841 (30.4)	1758 (32.7)
1	2567 (20.3)	1184 (22.0)
2	2262 (17.9)	979 (18.2)
3 or more	3981 (31.5)	1456 (27.1)
Prison_Offense (%)		
Drug	2681 (24.3)	936 (20.0)
Other	1302 (11.8)	611 (13.1)
Property	3742 (33.9)	2055 (44.0)
Violent/Non-Sex	2789 (25.3)	1008 (21.6)
Violent/Sex	522 (4.7)	61 (1.3)
Prison_Years (%)		
<1 yr	3666 (29.0)	1956 (36.4)
>3 yrs	3008 (23.8)	834 (15.5)

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

	false	true
1-2 yrs	3867 (30.6)	1762 (32.8)
2-3 yrs	2110 (16.7)	825 (15.3)
Prior_Arrest_Episodes_Felony (mean (SD))	5.43 (3.19)	6.49 (3.03)
Prior_Arrest_Episodes_Misd (mean (SD))	3.13 (2.29)	3.79 (2.21)
Prior_Arrest_Episodes_Violent (mean (SD))	0.98 (1.07)	1.10 (1.11)
Prior_Arrest_Episodes_Property (mean (SD))	2.04 (1.87)	2.66 (1.88)
Prior_Arrest_Episodes_Drug (mean (SD))	1.74 (1.69)	1.89 (1.70)
Prior_Arrest_Episodes_PPViolationCharges (mean (SD))	2.12 (1.92)	2.83 (1.92)
Prior_Arrest_Episodes_DVCharges (%)	1944 (15.4)	1041 (19.4)
Prior_Arrest_Episodes_GunCharges (%)	3213 (25.4)	1494 (27.8)
Prior_Conviction_Episodes_Felony (mean (SD))	1.36 (1.16)	1.56 (1.17)
Prior_Conviction_Episodes_Misd (mean (SD))	1.63 (1.53)	2.06 (1.56)
Prior_Conviction_Episodes_Viol (%)	3933 (31.1)	1919 (35.7)
Prior_Conviction_Episodes_Prop (mean (SD))	1.02 (1.15)	1.35 (1.21)
Prior_Conviction_Episodes_Drug (mean (SD))	0.75 (0.83)	0.81 (0.84)
Prior_Conviction_Episodes_PPViolationCharges (%)	3886 (30.7)	2041 (38.0)
Prior_Conviction_Episodes_DomesticViolenceCharges (%)	929 (7.3)	521 (9.7)
Prior_Conviction_Episodes_GunCharges (%)	1631 (12.9)	763 (14.2)
Prior_Revocations_Parole (%)	1079 (8.5)	625 (11.6)
Prior_Revocations_Probation (%)	1755 (13.9)	895 (16.6)
Violations_ElectronicMonitoring (%)	1038 (8.2)	315 (5.9)
Violations_Instruction (%)	2195 (17.4)	913 (17.0)
Violations_FailToReport (%)	959 (7.6)	383 (7.1)
Violations_MoveWithoutPermission (%)	1504 (11.9)	558 (10.4)
Delinquency_Reports (mean (SD))	1.07 (1.64)	0.79 (1.38)
Program_Attendances (mean (SD))	3.30 (3.87)	2.09 (3.18)
Program_UnexcusedAbsences (mean (SD))	0.41 (0.92)	0.41 (0.92)
Residence_Changes (mean (SD))	0.80 (1.01)	0.93 (1.07)
Avg_Days_per_DrugTest (mean (SD))	92.83 (115.38)	95.63 (123.25)
DrugTests_THC_Positive (mean (SD))	0.05 (0.12)	0.10 (0.18)
DrugTests_Cocaine_Positive (mean (SD))	0.01 (0.06)	0.02 (0.08)

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

	false	true
DrugTests_Meth_Positive (mean (SD))	0.01 (0.05)	0.02 (0.08)
DrugTests_Other_Positive (mean (SD))	0.01 (0.04)	0.01 (0.05)
Percent_Days_Employed (mean (SD))	0.56 (0.42)	0.28 (0.37)
Jobs_Per_Year (mean (SD))	0.82 (0.82)	0.64 (0.79)
Employment_Exempt (%)	1880 (14.9)	557 (10.4)
Recidivism_Arrest_Year1 (%)	0 (0.0)	5377 (100.0)

Imputation of missing values

Following exploratory data analysis, the `mice` package in R was used to create $m = 5$ imputed data sets to account for missing values in the data. All resulting model-based summaries, including model coefficients and standard error estimates, were then averaged over these five imputations using the appropriate methods in R. The proportion of missing values in the original data are detailed in the table below. Generally speaking, the more extensive the missingness, the less reliable the results of the imputation procedure. While 12% missing entries for gang affiliation was not bad, 24% missing for average days between drug tests was less than ideal.

Table 3: Proportion of missing values in NIJ data

	Prop Missing
Gang_Affiliated	0.12
Supervision_Risk_Score_First	0.02
Supervision_Level_First	0.07
Prison_Offense	0.13
Avg_Days_per_DrugTest	0.24
DrugTests_THC_Positive	0.20
DrugTests_Cocaine_Positive	0.20
DrugTests_Meth_Positive	0.20
DrugTests_Other_Positive	0.20
Percent_Days_Employed	0.02
Jobs_Per_Year	0.03

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Models

Type of model

We explored several different types of models for this task, including Bayesian and tree-based models. We ultimately settled on a standard logistic regression model for its simplicity and perceived good fit to this large ($n = 18,028$) data set.

Model selection

For the purpose of model development, we did not rely on any systematic backward or forward-type selection algorithm. Instead, we validated the entire model using likelihood ratio tests to compare full and reduced models. Once a candidate model had been selected based on this method, we examined measures of in-sample predictive performance. Once we were satisfied that the model was probably not over-fitting the training data, we generated predictions for the test set so that the model could be validated using out-of-sample performance metrics by the challenge administrators.

Our simplest logistic regression model for year one recidivism consisted of the following basic demographic and risk factor variables:

M_0 : Recidivism_Arrest_Year1 ~ Gender + Age_at_Release + Education_Level + Dependents + Gang_Affiliated + Supervision_Risk_Score_First

Our largest model proposal added criminal history, drug use, and employment variables:

M_1 : Recidivism_Arrest_Year1 ~ Gender + Age_at_Release + Education_Level + Dependents + Gang_Affiliated + Supervision_Risk_Score_First + Prison_Offense + Prison_Years + Prior_Arrest_Episodes_Felony + Prior_Arrest_Episodes_Misd + Prior_Conviction_Episodes_PPViolationCharges + Prior_Conviction_Episodes_Domestic ViolenceCharges + Prior_Conviction_Episodes_GunCharges + Prior_Conviction_Episodes_Prop + Prior_Conviction_Episodes_Viol + Prior_Revocations_Parole + Prior_Revocations_Probation + DrugTests_THC_Positive + DrugTests_Meth_Positive + DrugTests_Cocaine_Positive + DrugTests_Other_Positive + Percent_Days_Employed

In between the most complex and simplest models, we proposed another model. It was essentially a variation on M_1 which dropped the drug use and employment variables, since these variables were not included in the test data that was required to validate the model and finalize our submission. This model was:

M_2 : Recidivism_Arrest_Year1 ~ Gender + Age_at_Release + Education_Level + Dependents + Gang_Affiliated + Supervision_Risk_Score_First + Prison_Offense + Prison_Years +

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Prior_Arrest_Episodes_Felony + Prior_Arrest_Episodes_Misd + Prior_Conviction_Episodes_PPViolationCharges + Prior_Conviction_Episodes_Domestic ViolenceCharges + Prior_Conviction_Episodes_GunCharges + Prior_Conviction_Episodes_Prop + Prior_Conviction_Episodes_Viol + Prior_Revocations_Parole + Prior_Revocations_Probation

To compare the fitness of these models, we performed a series of likelihood ratio tests. Since the models were developed using multiply imputed training data, the package `mitml` was used instead of the standard base R `anova` function.

Model Comparison	F-stat	Pr(F>f)
M_0 vs. M_2	68.2	0.000
M_2 vs. M_1	250.9	0.000

Here we see that the most complex model, M_1 appeared to be the most fit model based upon the likelihood ratio test. That is, reducing the model from M_1 to M_2 or from M_2 to M_0 resulted in a significant decrease in model likelihood that was large enough to cause us to reject the null hypothesis that the additional coefficients in the more complex model were all zero. Based on this method, we would choose model M_1 as our final model and proceed with model validation and out-of-sample testing. Due to missing columns in the test data, we were forced to resort to using model M_2, since we were unable to generate predictions for the test data without the accompanying predictors for drug use and employment status.

Model validation

We chose to use the area under the receiver-operator characteristic (ROC) curve as the metric by which we judged in-sample classification performance. While the official metric of this competition was the Brier score, we did not wish to overfit to the method that the judges would be using to assess our model's out-of-sample performance. While ordinarily we may have used cross-validation to obtain an estimate of the model's out-of-sample performance, the complexities of working with multiply-imputed data made this a more challenging task. Instead, we cautiously consulted the in-sample model performance metrics and did our best to avoid the urge to overfit to the training data. Using the `ROCR` package, we calculated that the in-sample AUROC for model M_2 was 0.7 averaged over the five imputed data sets.

Final Model

Our final model was a logistic regression using formula M_2. A model summary is provided in the table below. We found that most of the coefficients were statistically significant at the 0.05 level based on an individual

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

t-test of each parameter, with the exception of some levels of certain categorical variables (e.g. dependents, education level, prior conviction on parole violation charges, prior conviction on gun charges, prior probation revocations). However, model selection using individual t-tests as a metric for whether a variable should be included is not generally recommended and was therefore not employed. Instead, we chose to make an evaluation of the overall model as described in previous sections.

Table 5: Summary of final model, including individual model coefficients, odds ratios, test statistics, and associated p-values.

Predictor	Coeff	Odds	SE	Statistic	p-value
(Intercept)	-2.01	0.13	0.10	-19.3	0.0000
GenderM	0.50	1.65	0.06	8.4	0.0000
Age_at_Release	-0.23	0.79	0.01	-16.6	0.0000
Education_LevelCollege	0.05	1.05	0.05	0.9	0.3883
Education_LevelHSgrad	0.10	1.10	0.04	2.6	0.0092
Dependents1	-0.03	0.97	0.05	-0.7	0.4792
Dependents2	-0.03	0.97	0.05	-0.6	0.5472
Dependents3 or more	-0.12	0.89	0.05	-2.5	0.0116
Gang_Affiliatedtrue	0.61	1.84	0.05	12.3	0.0000
Supervision_Risk_Score_First	0.04	1.04	0.01	4.1	0.0000
Prison_OffenseOther	0.10	1.10	0.06	1.5	0.1237
Prison_OffenseProperty	0.24	1.28	0.05	4.6	0.0000
Prison_OffenseViolent/Non-Sex	0.13	1.14	0.06	2.2	0.0297
Prison_OffenseViolent/Sex	-0.42	0.66	0.16	-2.6	0.0145
Prison_Years>3 yrs	-0.48	0.62	0.06	-8.1	0.0000
Prison_Years1-2 yrs	-0.23	0.80	0.04	-5.1	0.0000
Prison_Years2-3 yrs	-0.43	0.65	0.06	-7.6	0.0000
Prior_Arrest_Episodes_Felony	0.11	1.11	0.01	12.8	0.0000
Prior_Arrest_Episodes_Misd	0.09	1.10	0.01	9.5	0.0000
Prior_Conviction_Episodes_PPViolationChargestrue	0.00	1.00	0.04	0.0	0.9629
Prior_Conviction_Episodes_DomesticViolenceChargestrue	-0.05	0.95	0.07	-0.8	0.4156
Prior_Conviction_Episodes_GunChargestrue	-0.09	0.92	0.05	-1.6	0.1089
Prior_Conviction_Episodes_Prop	0.10	1.10	0.02	4.9	0.0000

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Predictor	Coeff	Odds	SE	Statistic	p-value
Prior_Conviction_Episodes_Violtrue	0.12	1.13	0.04	2.8	0.0045
Prior_Revocations_Paroletrue	0.24	1.28	0.06	4.2	0.0000
Prior_Revocations_Probationtrue	-0.08	0.93	0.05	-1.6	0.1094

In the above model summary, the odds ratio is the exponentiated value of the estimated model coefficient in the first column. For example, for gender, we found that male gender was associated with 65% higher odds of year one recidivism when all other predictors are held constant. A one-unit higher initial supervision risk score was associated with 4% higher odds of year one recidivism when all other predictors are held constant. Having three or more depends was associated with 11% lower odds of year one recidivism compared to no dependents when all other predictors are held constant. This is a small selection of the predictors included in this model, and these associations should in no way be considered causal.

Model comparison

As an alternative model to the logistic regression discussed above, we explored using a tree-based method, gradient boosted machines, using the `gbm` package in R. Tree-based methods are particularly useful when there may be many underlying interactions between variables, and they have an inherently different structure that linear models, making them a good benchmark by which to compare predictive performance.

Our GBM used the same predictors as logistic model M_2 above, and we considered 1,000 trees with depth up to 10, and a shrinkage term of 0.1. We found that this sample had an AUROC of 0.74. This was slightly higher than the in-sample AUROC for our logistic model. However, this alone was not a sufficient reason to choose the GBM model, since increased in-sample performance is often due simply to overfitting. Since the linear model seemed to perform well, our exploratory data analysis revealed that mostly linear relationships appeared to exist between the outcome of interest and the predictors, and there didn't appear to be many strong interactions between predictors (though we did not formally test this assertion since the number of possible two-way interactions among the 49 candidate predictors was formidable), we did not see this result as enough evidence to abandon the linear model in favor of this tree-based method.

Conclusions

Our final submission was a logistic regression model, identified in the *Models* section of this report as M_2, which used data about a parolee's demographics, general risk factors, criminal history, and prior parole

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

history to assess their probability of recidivating in their first year of release from prison.

Our goal during the model development process was to create the best overall model possible with respect to average accuracy. To that end, we made choices regarding the inclusion of predictors based upon the results of our exploratory data analysis, thoughts about algorithmic fairness when issues of race and other demographics were concerned, and the results of likelihood ratio test to compare the fitness of more complex models to those with only a subset of their predictors.

In order to develop the best possible model, we chose to include demographics like gender, but not to race. Gender seemed an appropriate predictor since the vast minority of parolees were women, and it was possible that their behaviors might differ from the larger male population of parolees. By contrast, including race as a predictor seemed like a recipe for introducing racial bias into our model, where the race variable might become a proxy for a host of socioeconomic disparities, which may not provide an objective assessment of an individual's probability of recidivating. The PUMA variable appeared to carry with it a similar risk, since sociological research and everyday experience indicate that many of America's communities continue to be de facto racially segregated (The Washington Post). Other variables like level of education may be within a parolee's control, assuming the institution where they were incarcerated offered educational opportunities. Other predictors related to their crime and sentence, including type of offense, number of years spent in prison, and prior arrests and convictions, also seemed relevant to assembling a coherent picture of the extent of the parolee's criminality. The relationships revealed during exploratory data analysis also largely supported that a linear relationship between these predictors and probability of first-year recidivism did exist.

With regard to the accuracy of the predictions for female parolees, our main focus for this subset of the population was to include gender and the type of offense that resulted in imprisonment. This was because in this sample, conditional on gender, women were more likely to commit property or drug offenses, and less likely to commit violent crimes. The distribution of prison sentences for females was also more heavily weighted toward lower values (i.e. less than three years) for women compared to the distribution of prison years for male parolees. So overall it was important to recognized the characteristics that distinguished the female parolee sample from the male parolee sample when building a model that would be sensitive to how gender is associated with the probability of first-year recidivism.

Finally, it is important to note that this model was developed with the goal of prediction in mind, from a data set that did not result from any designed experiment or prospective data collection. As such, the individual coefficients in the model should be viewed with caution. For example, the final model detailed above would indicate that prisoners with longer sentences (e.g. >3 years) were associated with lower odds

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

of first year recidivism relative to those with sentences of less than one year when all other predictors were held constant. While this coefficient may help contribute to an overall model with good average accuracy, determining whether estimates like this align well with currently accepted sociological research emanating from carefully designed, prospective studies is another important step towards validating a model if it were ever to be used for more than an academic exercise. We will also note that out-of-sample validation is the gold standard for any predictive model. While this model performed well compared to other submissions with regard to average accuracy using the Brier score, we did not have access to the response variables for the out-of-sample data to compute AUROC or other classification performance metrics. This is an important consideration in assessing the performance of a classifier.

Future considerations

For future work, we would recommend developing a version of this model that would take a careful approach to identify factors related to former prisoners' successful re-entry into society. While this data set provided some clues about associations between parolee characteristics and first-year recidivism, these are associations only and not causal relationships. They also come from a data set that was fairly limited in its scope in that it included data from just a few years in a single state. The data also excluded some individuals if they did not identify their race as black or white. Fixing the test data set, which was missing some variables, would also allow additional choices of final models that could be fully validated.

With regard to cutoff values for the predictions (i.e. NIJ's choice to use a value of 0.5), this did not affect our thinking during the development of this model. But it is worth noting that choosing this cutoff value is usually a component of the modeling process, and pre-specifying a value of 0.5 is probably sub-optimal. Finally, being penalized for false negatives as well as false positives may have changed our modeling strategy somewhat because it would require striking a more difficult balance in the model. This is an aspect of the modeling process where being able to fine-tune the cutoff value for predicting recidivism would be extremely useful.

Lastly, it may be worth considering whether viewing recidivism from year-to-year is the best approach for this type of modeling problem. Many well-developed methods exist for modeling time-to-event data, and it would be worthwhile to explore this framework for future modeling problems. In this way, separate categories would not be needed for years one through three, and a single model could be used to model all of the data.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

References

1. Brandon Greenwell, Bradley Boehmke, Jay Cunningham and GBM Developers (2020). gbm: Generalized Boosted Regression Models. R package version 2.1.8. https://CRAN.R-project.org/package=gbm

2. Stef van Buuren, Karin Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. URL https://www.jstatsoft.org/v45/i03/.

3. Simon Grund, Alexander Robitzsch and Oliver Luedtke (2021). mitml: Tools for Multiple Imputation in Multilevel Modeling. R package version 0.4-1. https://CRAN.R-project.org/package=mitml

4. Sing T, Sander O, Beerenwinkel N, Lengauer T (2005). "ROCR: visualizing classifier performance in R." *Bioinformatics*, *21*(20),

5. URL: http://rocr.bioinf.mpi-sb.mpg.de.

6. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

7. The Washington Post. *America is more diverse than ever - but still segregated*. Retrieved from: https://www.washingtonpost.com/graphics/2018/national/segregation-us-cities/

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.