



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

**Document Title: National Institute of Justice Recidivism
Challenge Report: Team Aurors**

Author(s): Team Aurors

Document Number: 305053

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.

National Institute of Justice Recidivism Challenge Report

Team: Aurors

Introduction

Forecasting algorithms have been used in a wide range of applications, from marketing purposes to credit applications. The fast advancement of forecasting techniques has made them more accurate and reliable, making them suitable for implementation in the criminal justice system. In the United States (US), law enforcement agencies have leveraged these techniques for predictive policing. According to Perry (2013), predictive policing consists of applying quantitative methods based on statistical forecasting to prevent crime, among other purposes. The National Institute of Justice (NIJ), following the priorities set by the US Department of Justice, is leading the research, development, and evaluation of methods to increase public safety and the fair administration of justice. For that purpose, the NIJ funded the Recidivism Forecasting Challenge, aiming to improve the ability to forecast recidivism.

The following report is the result of this challenge. This work uses individual and geographic-specific databases to design and calibrate forecasting models that help predict recidivism. The database was built using data from Georgia about people in parole supervision between 2013 and 2015, provided as a part of the NIJ Challenge, and data from the US Census Bureau about socio-economic characteristics at a geographic level. The forecasting models include regression analysis methods (binary logit and LASSO regressions) and machine learning techniques (random forest), combined through a model averaging procedure. The output consists of the percent likelihood of individuals recidivating within one, two, or three years from release. While this report explains the database construction process and the modeling approach, the results focus on the section of female paroles recidivating within three years, since that is the category where the team came in second place.

The report is structured as follows. Section 2 presents an overview of the relevant literature. Section 3 discusses the construction of the variables. Section 4 contains a discussion of the modeling process. Section 5 presents a summary of the findings. Section 6 focuses on future considerations. The report is finalized with conclusions and references.

Relevant Literature

This section provides a brief revision of the literature associated with prediction in criminology. Early research in the criminology field started with predicting the success or failure of parole (Burgess, 1928; Tibbitts, 1931). Using past records of the Illinois Parole System, Tibbitts (1931) predicted the observance or violation of parole with statistical methods. Lately, more sophisticated techniques have been used for prediction in criminology. Logistic regression (LR) was used to forecast recidivism in mentally-ill offenders (Gagliardi et al., 2004) and older adults (Rakes et al., 2018) released from prison.

Palocsay et al. (2000) predicted criminal recidivism using neural networks (NN). Through numerical experiments, they demonstrated that NN models have a higher classification accuracy for criminal recidivism compared with logistic regressions. The authors also highlighted the importance of selecting the network topology and the training methodologies in prediction models. Wadsworth et al. (2018) used an adversarially-trained NN model to predict recidivism and reduce racial bias. Wang et al. (2010) presented a general framework to predict criminal recidivism with support vector machines (SVM). The authors compared the performance of the NN, SVM, and LR models. As a result, in some cases, NN and SVM models outperform LR models; furthermore, combining the three models presents a higher performance than each single model. Tollenaar and Heijden (2012) questioned the best method to predict recidivism comparing statistical, machine learning, and data mining models. The authors demonstrated that traditional methods, e.g., LR, perform equally good or better than modern models.

Zeng et al. (2015), concerned about recidivism models that are accurate and easy to use and interpret, used a Supersparse Linear Integer Model (SLIM) that met these criteria. Ozkan (2017) predicted recidivism

in the U.S. criminal justice system through machine learning techniques. The author compared several machine learning models and found that XGBoost and neural networks outperformed the other models considered. Overall, the revision of the literature shows the continuing need to find methods that improve the overall performance and the reduction of bias of predictive models in criminology.

Variables

The dataset was constructed from two sources: (1) the data provided by NIJ, which consist of individual-specific variables of people in parole supervision between one to three years (this dataset from Georgia ranged from January 1st, 2013, to December 31st, 2015); and (2) data from **the Census Bureau's American Community Survey Public Use Microdata Sample (ACS PUMS)**. More specifically, we used geographic-specific data by PUMA area tied to the **25 regions (aggregated PUMAs) of the first dataset**. Among the variables from the ACS PUMS database, we selected those that **help better describe the socio-economic conditions of the PUMA area**, based on our judgment. The chosen variables were the following ones:

- Code of aggregated PUMAs in our dataset
- Average age
- Percentage of unemployment
- Average family income
- Percentage of households where grandparents are responsible for children
- Percentage of people with no health insurance coverage
- Percentage of houses with size less than 1 acre
- Percentage of houses with size less than 10 acres
- Percentage of commuters that do not use car
- Percentage of workers that work from home
- Average travel time of commute
- Percentage of households using food stamps
- Percentage of people born in a foreign country

- Income to poverty ratio
- Percentage of people with no public health insurance
- Average monthly rent
- Percentage of females
- Percentage of houses that are rented
- Percentage of houses that are not owned but the residents don't pay
- Average property value
- Percentage of households with zero cars, with one car, and with two cars

These variables do not include the base variables, e.g., “percentage of households with more than two cars” is not included since that would be the complement of the included variables on car ownership.

Acknowledging that there would be multicollinearity among the geographic-specific variables, since they did not have variation across individuals belonging to the same PUMA region, a **Principal Component Analysis (PCA)** was conducted to reduce the dimensionality of geographic-specific dataset while retaining the maximum amount of variability among them. **The first three components were used**, retaining together approximately 79% of the total variance, as can be seen in Figure 1.

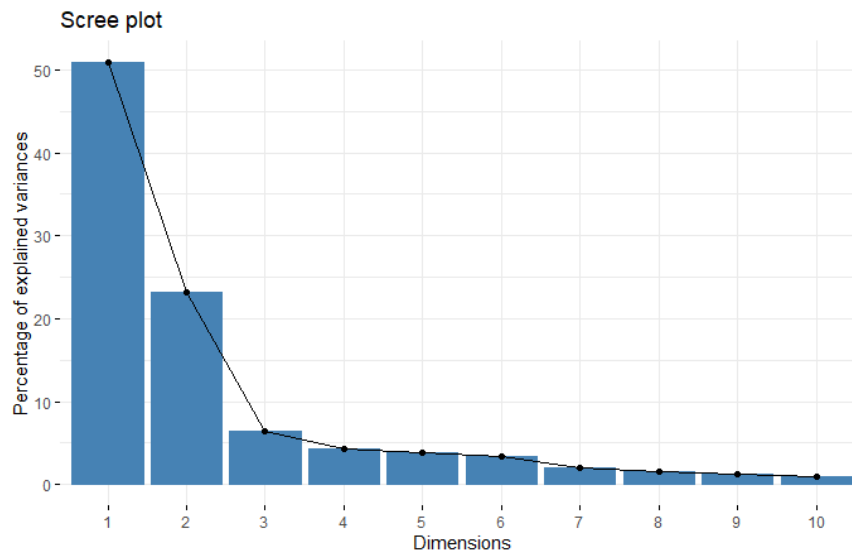


Figure 1. Percent of variance explained by each principal component (in decreasing order)

Regarding the data provided by NIJ, the authors converted the categorical variables into dummy variables. For the modeling process, the lowest category was used as the base; e.g., for the variable of age, the base case is the group of individuals between 18-22 years old.

Model

A brief description of the various model types used in the analysis is presented in this section. Particular attention is given to the results obtained after their application to predict the recidivism rates for females during the third year.

All models mentioned in this section were trained and tested to check their individual accuracy using an 80%-20% training-testing split of the training dataset available (without considering the testing dataset), i.e., 80% of the individual training records were randomly chosen to conform the training dataset and then used to estimate the different model structures. With the estimated models, a validation of their accuracy, separately or combined, was made using the remaining 20% of the data records.

Binary logit model

The authors used a binary logit model to forecast the recidivism for females within the third year. This model was very suitable for this scenario because it considers that the convict has the choice to come back to jail and predicts this probability using a logistic function. The specification of the model is shown in Equation 1 (Greene, 2003):

$$y_i^* = \beta' x_i + \varepsilon_i \tag{1}$$

Where y_i^* is the binary outcome variable of interest for individual i , x_i are the explanatory factors, e.g., the supervision and prison case information, prior Georgia criminal and community supervision history, the Georgia board of pardons and paroles conditions of supervision, and supervision activities. ε_i is the error term.

The logistic distribution is given by Equation 2, where $\Lambda(\cdot)$ is the cumulative distribution function (Berkson, 1944).

$$\text{Prob}(y_i = 1|x_i) = \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}} = \Lambda(x_i'\beta) \quad (2)$$

The binary logit model was applied with a step forward feature selection in R. The Akaike Information Criterion (AIC) of the model is 6876.2. Table 1 shows the results of the model.

Table 1: Binary Logit Model 3rd Year

Coefficients	Estimate	Std. Error	z value	Pr(> z)
Intercept	-1.6405	0.1350	-12.1550	< 2e-16 ***
prior_arrest_episodes_ppviolatio	0.0222	0.0248	0.8930	0.3718
gang	0.4491	0.0952	4.7190	0.0000 ***
age_48p	-1.3599	0.1555	-8.7460	< 2e-16 ***
prior_arrest_episodes_property	0.0576	0.0223	2.5770	0.0100 **
age_4347	-1.2092	0.1646	-7.3460	0.0000 ***
violations_instruction	0.2636	0.0825	3.1960	0.0014 **
prior_arrest_episodes_misd	0.0655	0.0231	2.8340	0.0046 **
female	-0.3164	0.0993	-3.1850	0.0014 **
educ_coll	-0.2102	0.0867	-2.4260	0.0153 *
age_3337	-0.8541	0.1456	-5.8650	0.0000 ***
age_3842	-0.8903	0.1553	-5.7350	0.0000 ***
prior_arrest_episodes_felony	0.0605	0.0174	3.4680	0.0005 ***
age_2832	-0.5901	0.1377	-4.2860	0.0000 ***
age_2327	-0.3279	0.1317	-2.4890	0.0128 *
prior_arrest_episodes_guncharges	0.1515	0.0705	2.1500	0.0316 *
condition_mh_sa	0.1275	0.0663	1.9210	0.0547 .
violations_movewithoutpermission	0.1716	0.0930	1.8450	0.0650 .
prior_conviction_episodes_misd	0.0636	0.0318	2.0030	0.0452 *
superv_sp	-0.1243	0.0723	-1.7190	0.0857 .
drugtests_thc_positive	0.4422	0.2754	1.6060	0.1083
depend_3	0.1146	0.0679	1.6870	0.0915 .
percent_days_employed	-0.2555	0.0922	-2.7720	0.0056 **
jobs_per_year	0.1206	0.0453	2.6600	0.0078 **
drugtests_meth_positive	0.9637	0.6264	1.5380	0.1240

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

LASSO regression

The LASSO (least absolute shrinkage and selection operator) was first introduced by Tibshirani, R. (1996). The LASSO regression is founded on the bias-variance tradeoff. As models become more complex,

their structure decreases bias, although the coefficient estimates suffer from high variance. Constraining the size of the coefficient estimates introduces bias but leads to substantial decreases in variance, hence, a significant reduction in the prediction error. The LASSO regression solves the optimization problem of Equation 3.

$$\min_{\beta} \sum_{i=1}^n (y_i - x_i' \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (3)$$

As observed in the previous equation, the LASSO regression uses a **tuning parameter λ that controls the strength of the penalty for the coefficients**. This penalty balances the idea of fitting a linear model and shrinking the coefficients. The nature of this penalty causes some coefficients to be shrunk to zero, thus performing variable selection in the linear model. Because of these reasons, the LASSO regression is very suitable for forecasting purposes when very complex (many variables) databases are involved. Besides, it provides a very competitive prediction error.

To ensure the LASSO regression reduces the prediction error, the tuning parameter λ must be chosen such that the mean squared error (MSE) is minimized. Therefore, it requires a **cross validation procedure**. We performed a K-fold cross validation with **1,000 iterations and $K = 10$** (Breiman and Spector, 1992 ; Kohavi, 1995). Figure 2 shows the result of the errors and standard error bands. The procedure yielded a $\lambda^* = 0.004924193$.

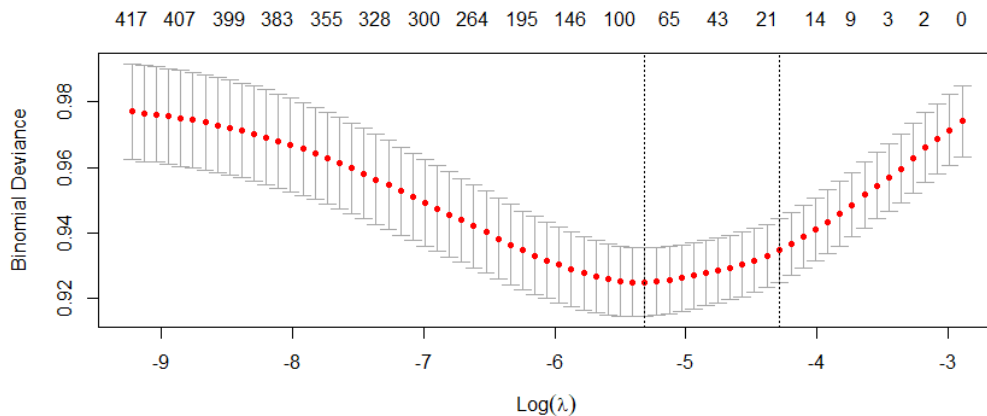


Figure 2. Errors and Standard Error Bands of 10-Fold Cross Validation

The LASSO regression was applied with λ^* in R, and the results are shown in Table 2, which contains the variables that were not shrunk to zero.

Table 2: LASSO Regression 3rd Year

Coefficient	Estimate	Coefficient	Estimate
Intercept	-1.9747	supervision_risk_score_first7:puma1	-0.0063
residence_puma8	-0.1353	supervision_risk_score_first6:puma2	0.0009
residence_puma10	0.0385	supervision_risk_score_first3:puma3	0.0062
residence_puma18	-0.2708	supervision_risk_score_first7:puma3	-0.0054
residence_puma20	0.0741	supervision_risk_score_first8:puma3	0.0731
residence_puma22	0.0428	prior_arrest_episodes_violent:black	0.0118
supervision_risk_score_first2	-0.0779	prior_arrest_episodes_drug:female	-0.0051
supervision_risk_score_first8	0.0034	prior_arrest_episodes_ppviolatio:puma2	0.0083
supervision_risk_score_first9	0.0514	prior_arrest_episodes_dvcharges:puma1	0.0137
prior_arrest_episodes_felony	0.0284	prior_arrest_episodes_guncharges:black	0.0191
prior_arrest_episodes_misd	0.0417	prior_conviction_episodes_drug:puma1	-0.0061
prior_arrest_episodes_violent	0.0031	prior_revocations_parole:black	0.0245
prior_arrest_episodes_property	0.0498	prior_revocations_probation:female	-0.1242
prior_arrest_episodes_ppviolatio	0.0355	condition_cog_ed:female	-0.1450
prior_arrest_episodes_guncharges	0.0774	violations_failto report:puma3	-0.0628
prior_conviction_episodes_misd	0.0499	violations_movewithoutpermission:puma1	0.0111
condition_mh_sa	0.0404	delinquency_reports:puma1	-0.0037
violations_instruction	0.2311	program_attendances:female	-0.0113
violations_movewithoutpermission	0.1084	avg_days_per_drugtest:puma2	0.0001
program_unexcusedabsences	0.0237	drugtests_thc_positive:puma1	-0.0006
drugtests_thc_positive	0.4276	drugtests_thc_positive:puma2	0.1182
drugtests_cocaine_positive	0.0968	drugtests_cocaine_positive:puma2	0.2796
drugtests_meth_positive	0.2505	drugtests_other_positive:black	-0.3222
percent_days_employed	-0.1032	jobs_per_year:puma1	0.0088
jobs_per_year	0.0123	employment_exempt:puma2	-0.0443
age_2327	0.1267	age_3337:puma2	-0.0016
age_3337	-0.0629	female:age_3842	-0.0931
age_4347	-0.2953	black:age_3842	-0.0964
age_48p	-0.4572	female:age_4347	-0.0633
gang	0.4778	gang:puma3	0.0321
superv_hi	0.0596	black:superv_hi	0.0334
educ_coll	-0.1721	superv_sp:puma1	-0.0060
yearp_3	-0.1045	female:depend_2	-0.0482
supervision_risk_score_first5:female	-0.2209	depend_2:puma2	0.0096
supervision_risk_score_first3:black	-0.1080	depend_3:puma2	-0.0425
supervision_risk_score_first4:black	-0.0168	yearp_12:puma3	0.0108
supervision_risk_score_first7:black	0.0029	female:yearp_23	-0.2011
supervision_risk_score_first8:black	0.1963	yearp_3:puma1	-0.0096
supervision_risk_score_first6:puma1	0.0255		

Random Forest model

A random forest model is based on the aggregation of several decision trees used to predict a given outcome of interest. In this case, the response variable, y_i , is a binary variable that shows if a given parolee

recidivates or not at a given year after release. Using the different features available, a decision tree corresponds to a partition of the features space in a way that each element of the partition is supposed to gather homogeneous individuals in terms of their response. Homogeneity is defined according to the nature of the response variable, with the mean squared error being the most popular choice for quantitative response variables, and the Gini index for qualitative outcomes (as in this case). Although decision trees are easily interpretable and more intuitive than other predictive models, they tend to suffer from **overfitting of the training dataset**, resulting in poor predictive ability. In that regard, Random Forest (RF) models were created to **aggregate the outcome of several decision trees** in a way that the entire random forest becomes more robust against overfitting and multicollinearity. The outcome of the various trees is **aggregated using a majority voting mechanism for qualitative variables**. The different trees in the forest are **constructed using bootstrapped samples** from the training portion of the dataset. To guarantee that the different trees considered provide uncorrelated information from each other, a **random selection of a subset of variables** according to which to split at each branch of the tree is considered. In this application, a total of **1,000 decision trees were aggregated** in the RF structure and **a choice of one out of eight randomly selected variables was implemented as part of the branching routine**. No variable selection mechanism was considered for this model, i.e., **all variables were kept in the final model**.

Model Averaging

With the aim of improving the overall predictive accuracy, a combination of the various models considered was constructed and tested against its individual components in terms of their predictive ability over the unused 20% testing portion. The resulting average model can be specified as:

$$\overline{M}_\alpha(x_i) = \alpha_1 M_1(x_i) + \alpha_2 M_2(x_i) + \dots + \alpha_n M_n(x_i). \quad (4)$$

$\overline{M}_\alpha(x_i)$ denotes the average model obtained as a convex combination of the models $M_1(x_i), M_2(x_i), \dots, M_n(x_i)$; and $\alpha_1, \alpha_2, \dots, \alpha_n$ are their corresponding coefficients that, in order to conform a convex combination, need to respect that $\alpha_i \geq 0, \forall i$ and

$$\alpha_1 + \alpha_2 + \dots + \alpha_n = 1. \tag{5}$$

The optimal set of coefficients was determined via the following optimization problem:

$$\vec{\alpha}^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*) = \arg \max \overline{BS}(\overline{M_\alpha}(\cdot)). \tag{6}$$

$\overline{BS}(\overline{M_\alpha}(\cdot))$ denotes the average Brier Score over the unused 20% testing portion.

Table 3 shows the predictive accuracy in terms of BS of the different models considered for year 3.

Table 3: Performance of the various models in terms of BS when predicting year 3

Binary logit model	LASSO regression	SVM model	RF model	Optimal average model*
0.14744	0.1478	0.151	0.1496	0.14724

* The optimal average model was a combination of the Binary logit model (with a coefficient of 0.75), the LASSO regression model (with a coefficient of 0.13) and the RF model (coefficient of 0.12)

Results

This section discusses the significance of variables for the regression analysis methods.

Binary logit model

The variables that were statistically significant at 0.1 level, and positive in the binary logit model are:

- Individuals verified by investigation as gang affiliated
- Any Violation for Not Following Instructions
- Any Violation for Moving Without Permission
- Any Prior GCIC Arrests with Gun Charges
- Parole Release Condition = Mental Health or Substance Abuse Programming
- Jobs Per Year While on Parole
- Dependents at Prison Entry = 3
- # Prior GCIC Arrests with Most Serious Charge = Misdemeanor
- # Prior GCIC Convictions with Most Serious Charge = Misdemeanor

- # Prior GCIC Arrests with Most Serious Charge = Felony
- # Prior GCIC Arrests with Most Serious Charge = Property

The variables that were statistically significant at 0.1 level and negative are:

- Ages Groups at time of prison release: 23-27, 28-32, 33-37, 38-42, 43-47, and 48 or more
- Female
- % Days Employed While on Parole
- Education Grade Level at Prison Entry At least some college
- First Parole Supervision Level Assignment = Specialized

The variables that were not significant in the binary logit model were:

- % Drug Tests Positive for Methamphetamine
- % Drug Tests Positive for THC/Marijuana
- # Prior GCIC Arrests with Probation/Parole Violation Charges

Considering that the Pr ($>|z|$) of these variables were in a range between 0.1083 and 0.3718, which is not far from the 0.1 significance level, the authors decided to leave them in the model. This decision was also supported by the fact that the purpose of the model was to predict recidivism.

LASSO regression

All statistically significant variables in the LASSO regression were shown in Table 2. The top ten with the highest positive relationship with recidivism, due to having the largest positive estimates, are:

- Individuals verified by investigation as gang affiliated
- % Drug Tests Positive for THC/Marijuana
- % Drug Tests Positive for Cocaine, in interaction with living in Puma 2
- % Drug Tests Positive for Methamphetamine
- Any Violation for Not Following Instructions

- First Parole Supervision Risk Assessment Score = 8, in interaction with race black
- Ages Groups at time of prison release: 23-27
- % Drug Tests Positive for THC/Marijuana, in interaction with living in Puma 2
- Any Violation for Moving Without Permission
- % Drug Tests Positive for Cocaine

The top ten variables with highest negative relationship (the most negative estimates) are:

- Ages Groups at time of prison release: 43-47 and +48
- % Drug Tests Positive for Other Drug, in interaction with race black
- Living in Puma 8 and 18
- First Parole Supervision Risk Assessment Score = 5, in interaction with female
- Years in Prison Prior to Parole Release = 2-3, in interaction female
- Education Grade Level at Prison Entry at least some college
- Parole Release Condition = Cognitive Skills or Education Programming, in interaction with female
- Any Prior Probation Revocations, in interaction with female

The significant variables have been discussed thoroughly above. The rest of the variables were not significant, and their coefficient shrunk to zero.

Additional models

The team also considered SVM model, using the variables that were discussed in the previous sections. We tried multiple kernels for SVM: linear, polynomial (up to 8th degree), and gaussian radial basis function (RBF). Overall, the model results were close to the other models when compared via testing the model on the 20% data records as shown in Table 3. Although SVM did not provide better insights when combined with the other models, its individual score was not far from the other models. Hence, SVM should be still considered among the models with potential in predicting recidivism.

Findings

Interestingly enough for the Binary logit model and the LASSO regression, the most potent variable that positively predicts recidivism is the indicator of percentage of drug tests positive for methamphetamine. This is the variable with the highest positive coefficient. The result indicates that the higher the percentage of drug tests positive for methamphetamine, the higher the probability of an individual returning to prison within three years after release. Other attributes that have a positive and strong association with recidivism were individuals verified by investigation as gang-affiliated, individuals with any violation for not following instructions, or any violation for moving without permission. The attributes related to prior Georgia criminal history, such as prior arrest or conviction, have a positive but lower influence on recidivism compared with the other variables.

Conversely, the categorical variables of age show a negative impact on the probability of recidivism. In essence, the older the person, the less likely to come back to jail within three years. These results are consistent with previous research that demonstrated that recidivism decreases with age (Rakes et al., 2018). Regarding gender, being female reduces the likelihood of recidivism. The variables of percentages of days employed while on parole, individuals with at least some college, and individuals with specialized first parole supervision level assignment, show a negative and lower association with recidivism.

In addition to the impact of the variables included as predictors, it is important to highlight the role of the model averaging technique in improving the predictive performance with respect to individual model components. A sensible combination of models developed under different paradigms could result in a performance boost for the resulting model as it was shown in this work.

Future Considerations

This section answers the NIJ questions regarding considerations for future efforts. To start with other evaluations metrics that could have been considered for this challenge, we consider that the Continuous

Ranked Probability Score (CRPS) could be a better indicator than the Brier Score (BS) for predictive accuracy of probability outcomes. This is because the former has proven to have more appealing properties.

With respect to the 0.5 threshold, it did not play a significant role in our case. However, we did notice that given the unbalanced proportion of recidivating parolees with respect to the non-recidivating ones biased our model to underpredict recidivism in some cases.

Related to the fairness penalty only considering false positives, it was a fact that did not affect our submission. We only measured the percentage of false positives when choosing the final model.

Regarding practical or applied findings that could help the field, we want to highlight that the findings about drug consumption being related to recidivism are enlightening in the sense that it can provide a clear direction towards which public policy should be focused. Also, the possibility of recording other information about the parolees during the three-year-period of observation might elicit other factors not captured in the variables already available.

With respect to ways to improve the challenge, we think that all provided information and guidelines were clear, and we do not think there was something specific that can be improved.

For next challenges, NIJ should disseminate the challenge announcement in broader circles to ensure reaching researchers that do not work in the field of the challenge topic. For example, we are group of engineers and although we were not familiar with the recidivism literature, we have good experience with modelling due to our work, that was the reason we joined. This broader dissemination might help participants from other fields to join and provide different insights on various aspects of the topic being discussed. Finally, it would be interesting to see different topics for next challenges.

Conclusions

Overall, the NIJ challenge was an intriguing competition that provided us with a better background on recidivism prediction along with more experience in building and combining predictive models. The team

started the challenge by conducting a literature review to have an informed decision about selecting suitable models for the problem at hand. We then started looking for variables that can provide additional insights on recidivism, and then conducted PCA to select the three components that can explain the most variance while decreasing the dimension of the dataset.

We then split the training dataset to 80% training and 20% testing, we used different models (Binary logit, LASSO regression, Random Forest, SVM) along with a model averaging technique to select the best predictive model using the 20% testing dataset. These models provided good insights with relatively small differences among them individually, ranging from a Brier score of 0.1474 for Binary logit to 0.151 for SVM. Combining these models via model averaging procedure provided better results with a Brier score of 0.14724. The final submitted model was a combination of Binary logit model (with a coefficient of 0.75), the LASSO regression model (with a coefficient of 0.13) and the RF model (coefficient of 0.12).

Looking at the Binary logit model and the LASSO regression, it was found that the indicator of percentage of drug tests positive for methamphetamine was the most prominent variable in positively predicting recidivism, while age was found to have an inverse relation with the probability to recidivate. These findings can provide decision makers with a clear idea about the target of their policies to reduce recidivism rates by targeting younger criminals with drug problems especially those with gang affiliation.

References

- Berkson, J. (1944). "Application of the Logistic Function to Bio-Assay." Journal of the American Statistical Association **39**(227): 357365.
- Burgess E.W. (1928). "Factors Determining Success or Failure on Parole." In: The workings of the indeterminate sentence law and the parole system in Illinois. Springfield, IL: Illinois State Board of Parole.

- Breiman, L., & Spector, P. (1992). Submodel Selection and Evaluation in Regression. The X-Random Case. *International Statistical Review / Revue Internationale De Statistique*, 60(3), 291-319. doi:10.2307/1403680
- Gagliardi, G.J., D. Lovell and P.D. Peterson (2004). "Forecasting Recidivism in Mentally Ill Offenders Released from Prison." *Law Human Behavior* 2: 133–155. <https://doi.org/10.1023/B:LAHU.0000022319.03637.45>
- Greene, W. H. (2003). Econometric Analysis, Pearson Education, Inc.
- Kohavi, R. (1995) A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Vol. 2, Montreal, 20-25 August 1995, 1137-1145.
- Ozkan, T. (2017). "Predicting Recidivism Through Machine Learning." PhD. Dissertation.
- Palocsay, S.W., P. Wang and R.G. Brookshire (2000). "Predicting Criminal Recidivism Using Neural Networks." *Socio-Economic Planning Sciences*, Elsevier 34(4): 271-284, December.
- Perry, W. L. (2013). "Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations". Rand Corporation.
- Rakes, S., S. Grace and S.J. Tripodi (2018). "Recidivism Among Older Adults: Correlates of Prison Re-entry." *Justice Policy Journal* 15:1-15
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267-288.
- Tibbitts, C. (1931). "Success or Failure on Parole Can be Predicted: A Study of the Records of 3,000 Youths Paroled from the Illinois State Reformatory.". *Journal of Criminal Law & Criminology*, 22: 11-50.
- Tollenaar, N. and P.G.M. Heijden (2012). "Which Method Predicts Recidivism Best? A Comparison of Statistical, Machine learning, and Data Mining Predictive Models." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 176: 565-584

- Wadsworth, C., F. Vera and C. Piech (2018). “Achieving Fairness through Adversarial Learning: an Application to Recidivism Prediction.” ArXiv: 1-5, abs/1807.00199.
- Wang, P., R. Mathieu, J. Ke and H. J. Cai (2010). "Predicting Criminal Recidivism with Support Vector Machine." 2010 International Conference on Management and Service Science: 1-9, doi: 10.1109/ICMSS.2010.5575352.
- Zeng, J., B. Ustun and C. Rudin (2015). “Interpretable Classification Models for Recidivism Prediction.” Journal of the Royal Statistical Society: Series A (Statistics in Society). 180. 10.1111/rssa.12227.