# Incapacitation, Recidivism and Predicting Behavior

Easha Anand
Intro. To Data Mining
Final Project

## I. Abstract

As the U.S. has shifted from a “just deserts” penal philosophy to a utilitarian one that prioritizes protecting society over punishing offenders, the problem of behavioral prediction and incapacitation has become of increasing importance. This project examines a database of 4,000 ex-offenders released from California state penitentiaries in the early 1960’s and re-interviewed in the late 1980’s, attempting to use the 68 variables in the database to predict recidivism. An initial multiple regression model had limited success, accounting for only a small percentage of the variance in the number of arrests to desistance, but a decision-tree algorithm more closely modeled ex-offender behavior. The white-box nature of the decision-tree algorithm also renders it a more suitable fit for a policy-related problem such as deciding whether or not to grant parole. Treating the number of arrests to desistance as a binary variable fared no better than treating the number of arrests to desistance as continuous.

## II. Background

In 2004, the total adult correctional population, both incarcerated and in the community, topped seven million.[1] As both the net number of offenders and the number of offenders who serve at least a portion of their sentence in the community has skyrocketed, the problem of deciding which offenders ought to be granted parole has taken on increasing importance. Ex-offenders are released into the community under one of three auspices: Probation, in which a sentence is suspended on judicial orders; parole, in which a local hearing examiner utilizes psychological data, data about institutional behavior and criminal career history factors to decide whether or not to have the offender be released under supervision; and good-time release, in which community transition is automatically triggered absent a certain number (or gravity) of infractions while incarcerated.[2] In all three instances, two major changes in penal philosophy and policy have caused dramatic shifts in the baseline numbers of ex-offenders released into the community and the baseline criterion for such release.

First, a shift in penal philosophy from retributive to utilitarian has by extension shifted the emphasis in such decision-making from normative to predictive. Parole’s original conception was utilitarian: Able-bodied British criminals were sold to colonial contractors under the assumption that work and contribution to society outweighed the deterrent effect of punishing such criminals. At the turn of the past century, when the federal Board of Parole was created, its ostensible focus, by contrast, was on rehabilitation: Community supervision was envisioned as part of an entire package that would reform criminals by treating underlying social, psychological and physical issues. However, ex-offender programs have been markedly unsuccessful at generating positive

---

[1] Glaze, Lauren and Seri Palla. “Probation and Parole in the United States, 2004.” *Office of Justice Programs, Department of Justice.* November 2004. Accessed 30 Apr. 2007. <http://www.ojp.usdoj.gov/bjs/pub/pdf/ppus04.pdf>

[2] Jacobson, Michael. “Downsizing Prisons: How to Reduce Crime and End Mass Incarceration.” *Ebrary.com*. Published 2005. Retrieved 30 Apr. 2006. <http://site.ebrary.com/lib/yale/Doc?id=10137182>

evidence for rehabilitation.[3] In the late 1970s and early 1980s, the "just deserts" model of punishment, which posited that the only factor that ought to enter into punitive decisions was the severity of the offense, was adopted in the Department of Justice; this was signified by the passage of the Comprehensive Crime Control Act of 1984, which established determinate sentencing, on the premise that each offense should be punished specifically and proportionally. Finally, in the past decade, there has been a resurgence of utilitarian policy-making at the federal and state levels, as manifested by the emergence of selective, rather than charge-based, incapacitation strategies. The result of this change is that the metric for success on parole is now quantitative—rate of recidivism—rather than qualitative (i.e. fairness, degree of rehabilitation, etc.).[4]

Second, a focus on parole as predictive has increased both the number of variables considered when making parole decisions and the number of guidelines constraining the scope parole officials' discretion. In 19 states, parole guidelines are entirely determinate: Information about the felon's original offense, behavior while incarcerated, prior convictions and psychological status are weighted accordingly to a set formula to determine whether or not the offender is eligible for parole.[5] Despite this renewed effort to inject objectivity into what has historically been a subjective process, however, there has been very little progress made in verifying the correlation between the data gathered and the likelihood of recidivism.

In this study, then, I attempt to model recidivism rates based on 68 variables assessed for 5,000 offenders in the California penal system using two techniques from this class: Multiple regression and decision trees. I treat the number of arrests to desistance both as a continuous and as a binary variable.

## III. The Dataset

### *A. Characteristics*

The sample studied examined data on 6,000 men released from the California penal system in the early 1960's. Data collected include life history information, responses to inmate questionnaires, psychological data, institutional records and criminal career information; a follow-up survey in 1988 managed to locate 4,987 of these individuals and collected data about intervening arrests and convictions.[6]

Each crime is weighted along six dimensions: Nuisance, physical harm, property damage, drugs, fraud and crimes against social order. Over half the crimes committed after release—54.4 percent—fall into the category of entirely "nuisance" crimes; however, over 10,000 serious crimes (including 184 homicides, 2,084 assaults, 126 kidnappings,

---

[3] Farabee, David. *Rethinking Rehabilitation: Why Can't We Reform our Criminals?*. Washington, D.C.: AEI Press, 2005.
[4] Stith, Kate and Jose Cabranes. *Fear of Judging*. Chicago: University of Chicago Press, 1998.
[5] *Ash v. Reilly*, 431 F.3d 826 (D.D.C. 2006).
[6] For a full list of variables considered, see codebook at http://www.icpsr.umich.edu/cgi-bin/bob/file?comp=none&study=9922&ds=1&file_id=681904.

144 rapes, 2,756 burglaries and 1,193 robberies) have also been committed by the sample population.

*B. Class Variable*

In order to limit the scope of the project, I chose to attempt prediction of only one variable, the number of arrests to desistance.

On average, the men in question were arrested six times over the course of the 20 years following their release; among only those ex-offenders who were re-incarcerated, the average number of arrests was 8.5. The median number of re-incarcerations was 1.68, though almost one in three men was never reincarcerated.

There are three primary issues with selecting this variable as the class variable. First, two features of the California penal system's recordkeeping protocols insert bias into the sample. Crimes committed across state lines are not recorded within California (and, in fact, do not enter into parole determinations). Also, if an ex-offender remains law-abiding for two decades and is over the age of 60, his record is purged from the system. The latter in particular introduces a decided skew, as such ex-offenders are, in general, the type that parole boards seek to identify.

Second, regressing on the number of arrests—rather than the number of convictions or incarcerations—allows factors about local police force activity to affect recidivism data. However, I believe that using convictions to desistance or incarcerations to desistance introduces additional unnecessary variables, namely the different burdens required for the different crimes and the leniency of the sentencing judge in question.

Finally, though I posit that ex-offenders who remain arrest-free were more suitable for parole to begin with, the causality may be reversed. That is, it may be variables relating to the parole program itself that determines whether ex-offenders are deterred, in which case this analysis is moot.

*C. Pre-processing*

First, individuals for whom arrest-to-desistance rates were not available were filtered from the sample. Second, four variables—prior periods of incarceration; number of arrest free periods; type of offense; and type of offense—were turned into class variables to prevent skewing based on an arbitrary numbering system. In addition, the dataset was randomly divided into a 3,000-member training set and a 1,987-member test set.

## IV. Analysis

*A. Multiple Regression*

An initial attempt to model the arrest-to-desistance rate via multiple linear regression. Though the model successfully accounted for 16 percent of the variance in the general case, it only accounted for 6 percent of the variance in the case of violent crimes, presumably the more relevant gauge of success.

Table 1: Number of Arrests to Desistance

| Predictor | Regression Coeff | Standardized Reg. Coeff | T |
|---|---|---|---|
| Priors | 1.115 | .270 | 11.02 |
| Age | -.104 | -.144 | -6.39 |
| Drugs | -2.155 | -.154 | -7.94 |
| Serious | -.015 | -.058 | -2.92 |
| Free | -.899 | -.062 | -3.18 |
| PriorsP | -.413 | -.085 | -2.37 |
| Type | -.706 | -.05 | -2.31 |
| Alias | .343 | .046 | 2.31 |

**Notes:**

- $p<.05$
- $R^2 = .159$
- PriorsP = priors that resulted in prison incarceration; Free = number of arrest-free periods; Drugs = binary measure of whether original offense was drug-related; Age = age at first incarceration; Serious = code number of most serious crime.
- Only those variables with a statistically significant correlation to number of arrests to desistance were used.

Table 2: Number of Arrests to Desistance, Violent Crimes Only

| Predictor | Regression Coeff | Standardized Reg. Coeff | T |
|---|---|---|---|
| Priors | -.022 | -.174 | -7.85 |
| Age | .134 | .184 | 7.45 |
| InstP | .253 | .076 | 3.35 |
| PriorsP | -.066 | -.077 | -2.91 |

**Notes:**

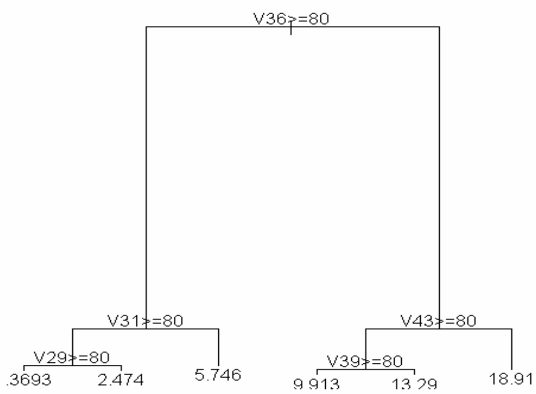
- $p<.05$
- $R^2 = .061$
- InstP = whether or not the original offense was a crime against a person.
- N = 1,998
- Only those variables with a statistically significant correlation to number of arrests to desistance were used.

*B. Decision Tree*

Given both the regression model's lack of success in predicting the most important instances of recidivism (violent crimes) and the "black-box" nature of regression models, I opted for an alternative algorithm. The primary advantage of the decision-tree—its ease of use and intelligibility, as well as the intuitive nature of the model—is particularly critical given the policy applications in question: For inmates to achieve some degree of certainty about their release date, a model must facilitate easy use by non-experts. To execute this algorithm, the Recursive Partitioning library was utilized to create regression trees designed to predict the arrests to desistance.[7]

An initial attempt, using all available variables, split first on Most Serious Charge at 5th Arrest Episode (V36), then on Most Serious Charge at 1st Arrest Episode (V31) and Most Serious Charge at 12th Arrest Episode (V43), and finally on Actual Time Incarcerated (V29) and Most Serious Charge at 8th Arrest Episode (V39). With the exception of the Actual Time Incarcerated variable, the remaining variables all relate to criminal career factors; although such measures are undoubtedly relevant, it would appear that the particular ordering (with 5th episode giving the most additional information about likelihood of recidivism) represents a situational artifact.

Figure 1: Decision Tree, with Type = Arrests to Desistance




[7] Authors Terry M Therneau and Beth Atkinson atkinson@mayo.edu, R port by Brian Ripley <ripley@stats.ox.ac.uk>.

However, despite the seemingly arbitrary nature of the root node variable, the model achieves a relatively high degree of accuracy with regards to broad classification categories. This suggests that its primary use may be to assess the relative risk values of paroling various criminals. Later, I attempt to distinguish between the 0 arrest-to-desistance rate (i.e. unlikely to offend again) and all other arrest-to-desistance rates.

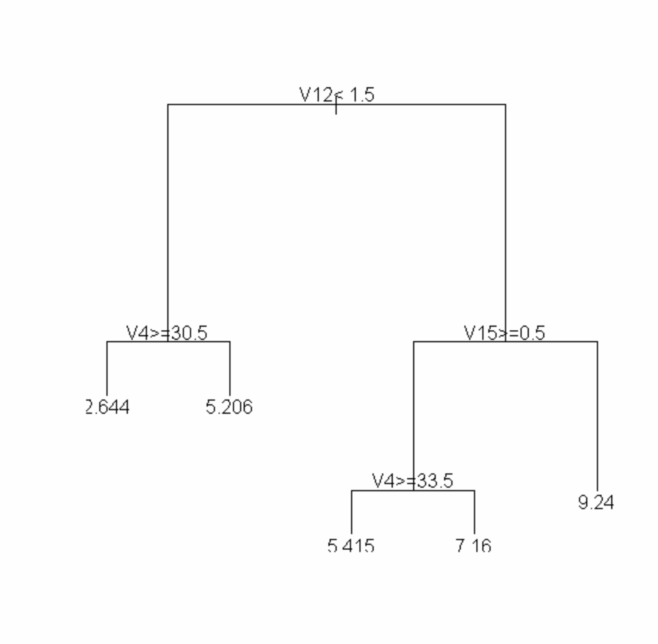
Table 3: Predicted vs. Actual Values of Arrest to Desistance

| | true | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pred | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 0.369318181818182 | 889 | 518 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2.47368421052632 | 0 | 0 | 430 | 387 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5.7457898957498 | 0 | 0 | 0 | 0 | 307 | 295 | 248 | 202 | 195 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9.91324200913242 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 167 | 142 | 129 | 0 | 0 | 0 |
| 13.287841191067 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 132 | 99 | 92 |
| 18.9109589041096 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | true | | | | | |
|---|---|---|---|---|---|---|
| pred | 15 | 16 | 17 | 18 | 19 | 20 |
| 0.369318181818182 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2.47368421052632 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5.7457898957498 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9.91324200913242 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13.287841191067 | 79 | 0 | 0 | 0 | 0 | 0 |
| 18.9109589041096 | 0 | 73 | 63 | 55 | 45 | 348 |

Upon limiting the set of dependent variables to offender-related variables only, the regression tree created splits first on Prior Periods of Incarceration (V12), then on Age When First Released (V4) and History of Opiate Use (V15).

Figure 2: Decision Tree, with Type = Arrests to Desistance; Offender-Related Variables Only




However, its predictive power is lower than the predictive power of the model that utilizes all available dependent variables. In particular, the model successfully classifies the 0 arrest-to-desistance rate offenders—those who would be the most viable parole candidates—successfully into the lowest category 30 percent of the time.

Table 4: Predicted vs. Actual Values of Arrest to Desistance; Offender-Related Variables Only

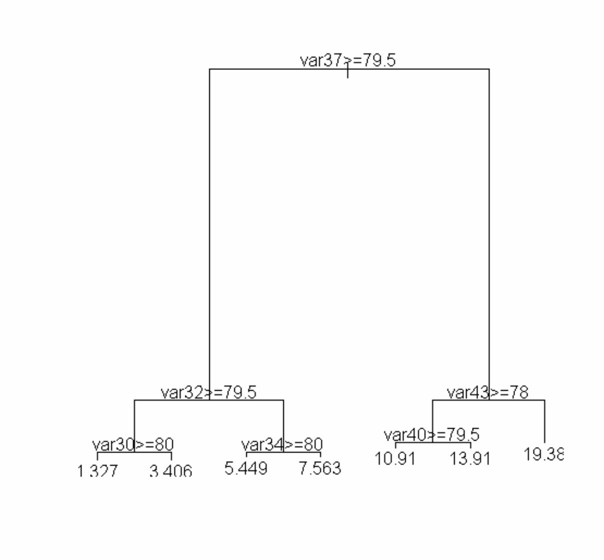
| pred \ true | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2.6441893830703 | 264 | 126 | 60 | 70 | 31 | 34 | 33 | 10 | 12 | 11 | 7 | 6 | 9 | 5 | 3 |
| 5.20592383638928 | 165 | 80 | 71 | 53 | 40 | 39 | 39 | 28 | 31 | 24 | 15 | 19 | 15 | 20 | 6 |
| 5.41496062992126 | 240 | 150 | 122 | 118 | 106 | 82 | 59 | 53 | 51 | 39 | 27 | 28 | 30 | 19 | 21 |
| 7.15983263598326 | 141 | 95 | 115 | 95 | 77 | 86 | 63 | 65 | 50 | 44 | 41 | 34 | 42 | 23 | 27 |
| 9.23976608187134 | 79 | 67 | 63 | 51 | 53 | 54 | 54 | 46 | 51 | 49 | 52 | 43 | 36 | 32 | 35 |

| pred \ true | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|
| 2.6441893830703 | 2 | 4 | 5 | 0 | 0 | 5 |
| 5.20592383638928 | 9 | 7 | 5 | 5 | 4 | 34 |
| 5.41496062992126 | 11 | 13 | 7 | 14 | 12 | 68 |
| 7.15983263598326 | 30 | 20 | 19 | 15 | 9 | 104 |
| 9.23976608187134 | 27 | 29 | 27 | 21 | 20 | 137 |

To test the salience of the decision-tree model with regards to the most serious offenses (i.e. violent crimes), I used the rpart algorithm on a subset of the database, dealing only with those who, upon release, committed crimes that scored high on the "crimes against persons" dimension. The resulting decision tree again relied primarily on particular episodes within the criminal history, splitting first on Most Serious Charge at 6$^{th}$ Arrest

Episode (V37), then on Most Serious Charge at 2nd Arrest Episode (V32) and Most Serious Charge at 12th Episode (V43), then on Most Serious Charge at 3rd Episode (V34), Most Serious Charge at 9th Episode (V40) and Family Criminal History Score (V30).

Figure 3: Decision Tree, with Type = Arrests to Desistance; Violent Offenses Only



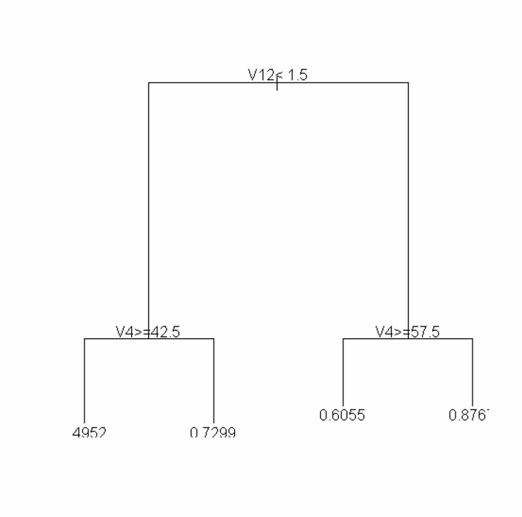

Hearteningly, within the smaller sample, the model performs relatively well at categorizing potential parolees into broad risk categories. In particular, it successfully categorizes all of the lowest-risk parole candidates into the lowest-risk category.

Table 5: Predicted vs. Actual Values of Arrest to Desistance; Violent Offenses Only

| true | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pred | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 17 | 18 | 19 | 20 |
| 1.32743362831858 | 3 | 70 | 40 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3.40579710144928 | 0 | 0 | 0 | 41 | 28 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5.44897959183673 | 0 | 0 | 0 | 0 | 0 | 27 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7.5625 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 25 | 19 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10.90625 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 7 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13.9090909090909 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 8 | 6 | 0 | 0 | 0 | 0 |
| 19.3793103448276 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 4 | 4 | 19 |

It is an open question where false positive or false negatives are most relevant in assessing parole risk. On the one hand, had all of the felons in this dataset remained locked up, 30,000 arrests could have been prevented. On the other, had all of the felons in this dataset remained locked up, 1,413 people would have remained unnecessarily locked up. I assess the likelihood of risk-free parole—that is, treat the “arrest-to-desistance” variable as binary, assigning it a value of 0 if the ex-felon never re-offended and a value of 1 otherwise—and examine the power of the decision tree generated. The root node of the tree in question is Prior Periods of Incarceration (V12), followed by Approximate Age When Released (V4).

Figure 4: Decision Tree, with Type = Arrests to Desistance (Binary)




However, the model demonstrates little to no predictive accuracy.

Table 3: Predicted vs. Actual Values of Arrest to Desistance (Binary)

| | true | |
|---|---|---|
| pred | 0 | 1 |
| 0.495238095238095 | 106 | 104 |
| 0.605504587155963 | 43 | 66 |
| 0.729933110367893 | 323 | 873 |
| 0.876700177409817 | 417 | 2965 |

**V. Acknowledgements**

- Trevor Gardner, University of California, Berkeley, for providing the dataset;
- Inter-University Consortium for Political and Social Research, for providing the codebook; and
- Jiang Du, for suggesting the decision tree algorithm.