Predicting Recidivism to Reimagine Incarceration

By: Sierra Sikorski

1.9 million people are incarcerated in the United States

55 thousand

people are sentenced to life in prison without the possibility of parole

1 out of 2

people released return to prison

3 out of 100

people who get a bachelor's degree while in prison return

What factors account for successful reentry after incarceration?

Data Source & Features

Source: Office of Justice Programs

Last updated in 2021

Contained: just over 50 variables and almost 26,000 observations

Strength: Many interesting variables

if they have changed residences, how many jobs they've had, and percentage days employed

Weakness: Lack of some proven relevant variables

No variable for if someone moves from where they were arrested.

General Approach

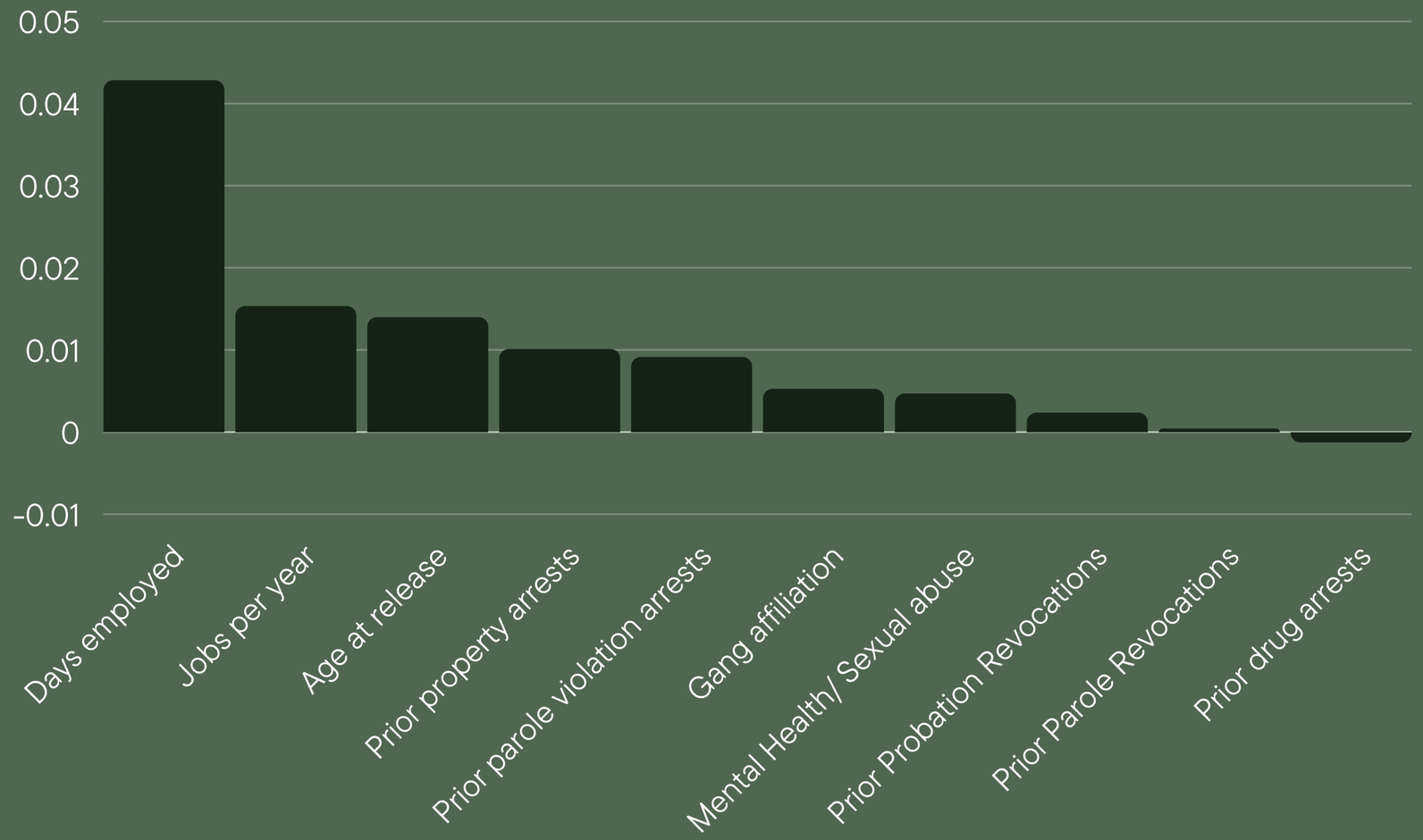
- Running two models: Linear Regression and Support Vector Machine
- Variables are scaled
- Each model is run twice.
    - Once with handpicking the included variables
    - Once with the sci-kit learn feature selector picking 10 variables
- For each model, I will look at the overall accuracy score, the confusion matrix, and the recall score.

Model Evaluation

	HAND PICKED: LOGISTIC REGRESSION	AUTO-SELECTED: LOGISTIC REGRESSION	HAND PICKED: SVM	AUTO-SELECTED: SVM
Accuracy	68%	71.2%	69.7%	72.3%
Recall	80.3%	81%	83.7%	86.2%

Feature Importance

For Auto Selected Models



Conclusions

For this data SVM tended to work the best

Different than literature review

Recall scores varied much more than accuracy scores

Models got better at inferring who returned to prison

Percent Days employed was most significant variable

Legislation like ban the box has merit

References

Ashley Nellis, P. D. (2022, November 11). Nothing but time: Elderly Americans serving life without parole. The Sentencing Project. Retrieved April 23, 2023, from https://www.sentencingproject.org/reports/nothing-but-time-elderly-americans-serving-life-without-parole/

Nij's Recidivism Challenge Full Dataset: Office of Justice Programs. NIJ's Recidivism Challenge Full Dataset | Office of Justice Programs. (2021, July 15). Retrieved March 3, 2023, from https://data.ojp.usdoj.gov/Courts/NIJ-s-Recidivism-Challenge-Full-Dataset/ynf5-u8nk

Recidivism of federal offenders released in 2010. United States Sentencing Commission. (2022, May 9). Retrieved March 2, 2023, from https://www.ussc.gov/research/research-reports/recidivism-federal-offenders-released-2010

Spivey-Jones, R. (2022, February 7). Written Testimony in Support of Restoring TAP Funding to Incarcerated Individuals. NY Assembly. Retrieved March 2, 2023, from https://nyassembly.gov/write/upload/publichearing/001273/003778.pdf

Tollenaar, N., & van der Heijden, P. G. (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(2), 565–584. https://doi.org/10.1111/j.1467-985x.2012.01056.x

Wagner, W. S. and P. (n.d.). Mass incarceration: The whole pie 2023. Prison Policy Initiative. Retrieved April 23, 2023, from

Appendix

Choosing the number of columns

	FEATURE COUNT 2	FEATURE COUNT 5	FEATURE COUNT 10	FEATURE COUNT 20	FEATURE COUNT 30	FEATURE COUNT 40
SVC	Recall: 80% Accuracy: 64.5%	Recall: 84.2% Accuracy: 70.5%	Recall: 83.5% Accuracy: 71.8%	Recall: 83.6% Accuracy: 71.8%	Recall: 82.6% Accuracy: 71.8%	Recall: 82.2% Accuracy: 72%
Logistic	Recall: 78.8% Accuracy: 64.4%	Recall: 79.4% Accuracy: 70.3%	Recall: 80.6% Accuracy: 71.7%	Recall: 81% Accuracy: 72%	Recall: 81% Accuracy: 72.1%	Recall: 80.7% Accuracy: 72.1%

Hand-Picked Variables

- **Gender**
- Dependents
- **Prior Parole or Probation Revocations**
- **Age at release**
- Supervision level
- Education level
- Program attendance
- **Percent days employed**
- Prior misdemeanor or felony arrests
- Prison Years

Auto-Selected Variables

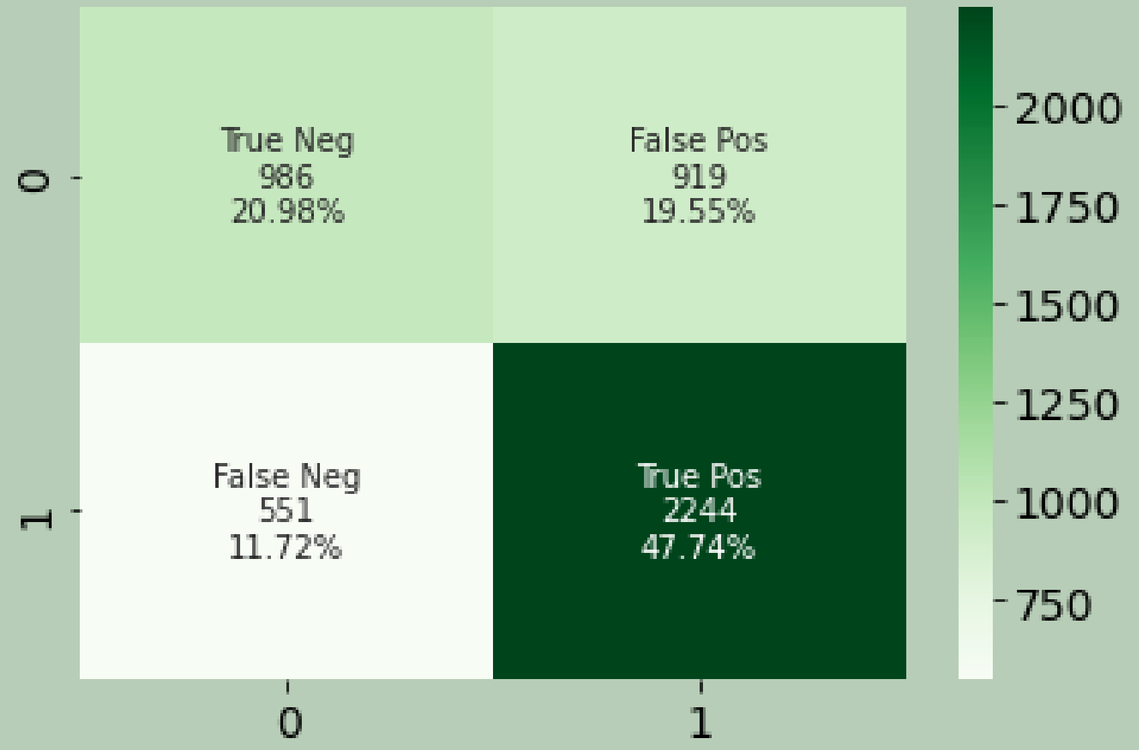
- **Gender**
- **Percent days employed**
- Jobs per year
- Gang affiliation
- **Age at release**
- Prior property arrests
- Mental health/ sexual abuse
- **Prior Parole or Probation Revocations**
- Prior drug arrests

Handpicked: Logistic Regression

Accuracy Score: 68.7%

Recall: 80.3%

Correctly identified 80% of true recidivism cases.

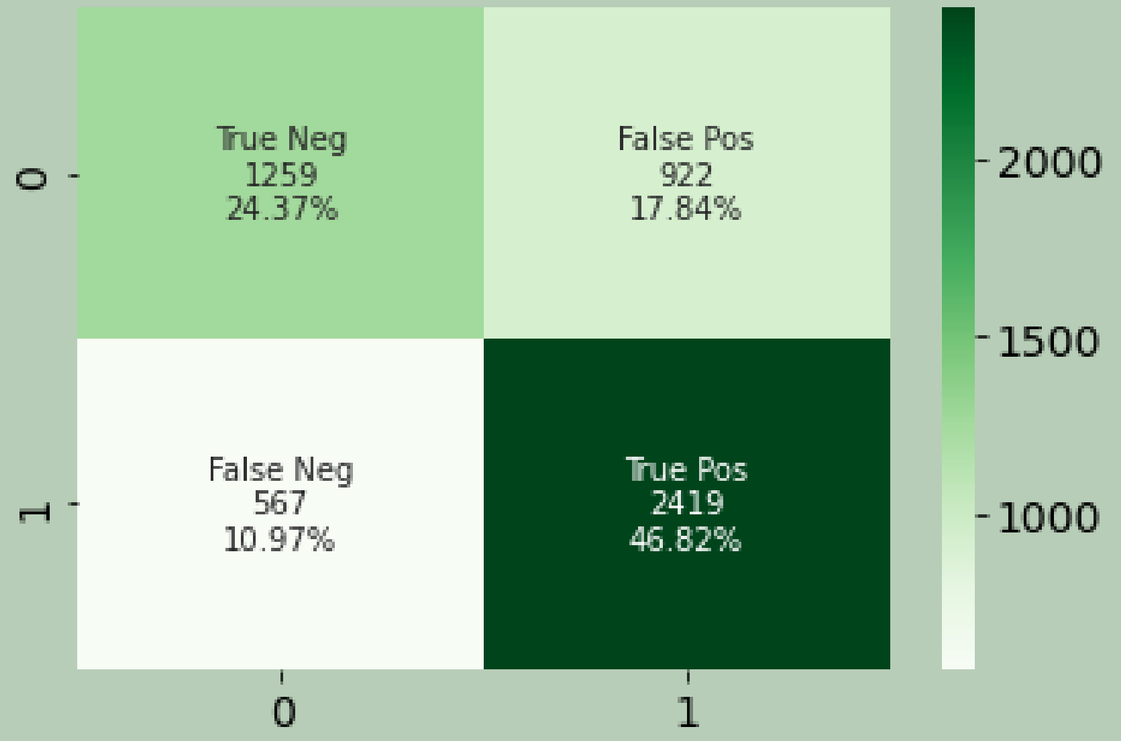


Auto-Selection: Logistic Regression

Accuracy Score: 71.2%

Recall Score: 81%

Correctly identified 81% of true recidivism cases.

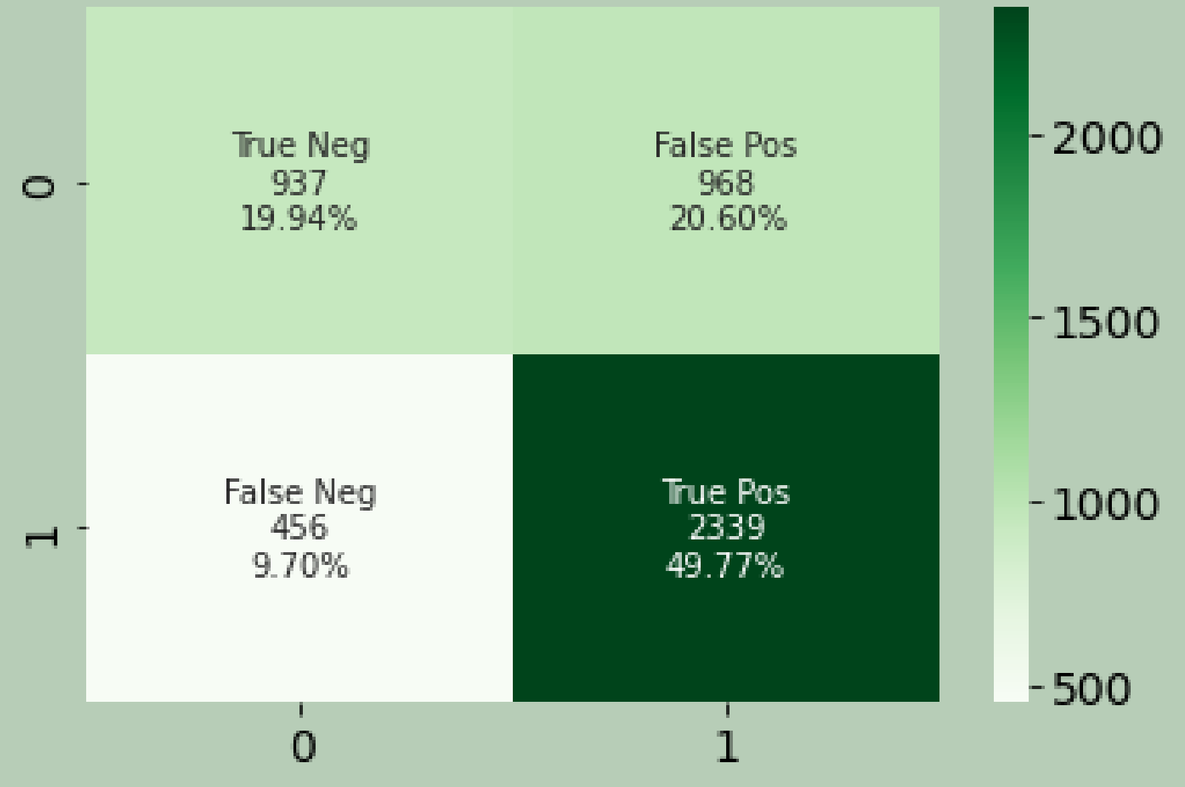


Handpicked: Support Vector Machine

Accuracy Score: 69.7%

Recall Score: 83.7%

Correctly identified almost 84% of true recidivism cases.



Auto-Selection: Support Vector Machine

Accuracy Score: 72.3%

Recall Score: 86.2%

Correctly identified 86.2% of true recidivism cases.