Predicting the Probability of Recidivism

By: Kyle Hanson, Christine Orosco, Myra Rust

Executive Summary

This report provides analysis and evaluation of machine learning techniques used to predict the probability of recidivism. Methods of analysis include leveraging historical data, machine learning, and predictive modeling to determine the probability of a convict to reoffend. Multiple models were evaluated, and performance was measured on accuracy rates. The best performing model achieved an accuracy rate of 66%. This result is good, but the gravity of decisions made on personal freedoms requires a higher accuracy to implement the model for use. It is recommended that further information is collected and aggregated to the current dataset in attempt to increase model accuracy.

Introduction/Background

Again, and again, you read news articles or watch television programs about persons who were previously incarcerated, committing additional criminal acts after being released from prison. The question that we ask ourselves is, could these additional crimes have been prevented? Ideally, if we could determine who is likely to reoffend, better decisions can be made on initial sentence lengths or if a convict should be released on parole.

According to a 2005 study conducted by the Bureau of Justice Statistics, "67.8% of released prisoners were arrested for a new crime within 3 years." (Durose et al., 2014). The reality is recidivism rates are significant and that must weigh heavily on the minds of judges and parole board officials while making decisions about a person's fate. As it stands today, decision makers are provided with high-level statistics on recidivism rates, but can we do better? Can we provide them with more accurate information about the likelihood that an individual would reoffend? By doing this, it would allow judges and parole board officials to make better decisions, not based on generalizations, but on individualism and hopefully mitigate or prevent additional crime. This report contains all steps that were taken to identify the most accurate model possible, including data cleaning, transforming, model selection, model evaluation, results, and recommendations.

Preliminary Analysis

Data Source: The 3-Year Recidivism for Offenders Released from Prison in Iowa dataset contains data collected from 2010-2018 by the Iowa Department of Corrections. The dataset consists of 26,020 observations and is predominately categorical variables of demographic and crime related information. The target variable is binary and indicates whether a convict reoffended within the three-year period.

	Fiscal Year Released	Recidivism Reporting Year	Main Supervising District	Release Type	Race - Ethnicity	Age At Release	Sex	Offense Classification	Offense Type	Offense Subtype	Return to Prison	Days to Return	Recidivism Type	New Offense Classification	New Offense Type	New Offense Sub Type	Target Population
0	2010	2013	7JD	Parole	Black - Non-Hispanic	25-34	Male	C Felony	Violent	Robbery	Yes	433.0	New	C Felony	Drug	Trafficking	Yes
1	2010	2013	NaN	Discharged – End of Sentence	White - Non-Hispanic	25-34	Male	D Felony	Property	Theft	Yes	453.0	Tech	NaN	NaN	NaN	No
2	2010	2013	5JD	Parole	White - Non-Hispanic	35-44	Male	B Felony	Drug	Trafficking	Yes	832.0	Tech	NaN	NaN	NaN	Yes
3	2010	2013	6JD	Parole	White - Non-Hispanic	25-34	Male	B Felony	Other	Other Criminal	No	NaN	No Recidivism	NaN	NaN	NaN	Yes
4	2010	2013	NaN	Discharged – End of Sentence	Black - Non-Hispanic	35-44	Male	D Felony	Violent	Assault	Yes	116.0	Tech	NaN	NaN	NaN	No

Figure 1: Original Dataset

Data Cleaning:

- Future crime related information would be unknown at the time of prediction and was therefore removed to eliminate introducing bias into models.
- Race/ethnicity information was removed to mitigate propagating racial bias.
- Logistical and other variables that provided no predictive value were removed.
- Inconsistences in the method of data collection and reporting were noted and accounted for by combining like variables.
- Missing values were present for the Age at Release and Sex variables and were imputed to the modes of those variables.

	year	release	age	sex	classification	type	subtype	reoffend
0	2010	Parole Granted	25-34	Male	C Felony	Violent	Robbery	1
1	2010	Discharged - Expiration of Sentence	25-34	Male	D Felony	Property	Theft	1
2	2010	Parole Granted	35-44	Male	B Felony	Drug	Trafficking	1
3	2010	Parole Granted	25-34	Male	B Felony	Other	Other Criminal	0
4	2010	Discharged - Expiration of Sentence	35-44	Male	D Felony	Violent	Assault	1

Figure 2: Dataset post Cleaning

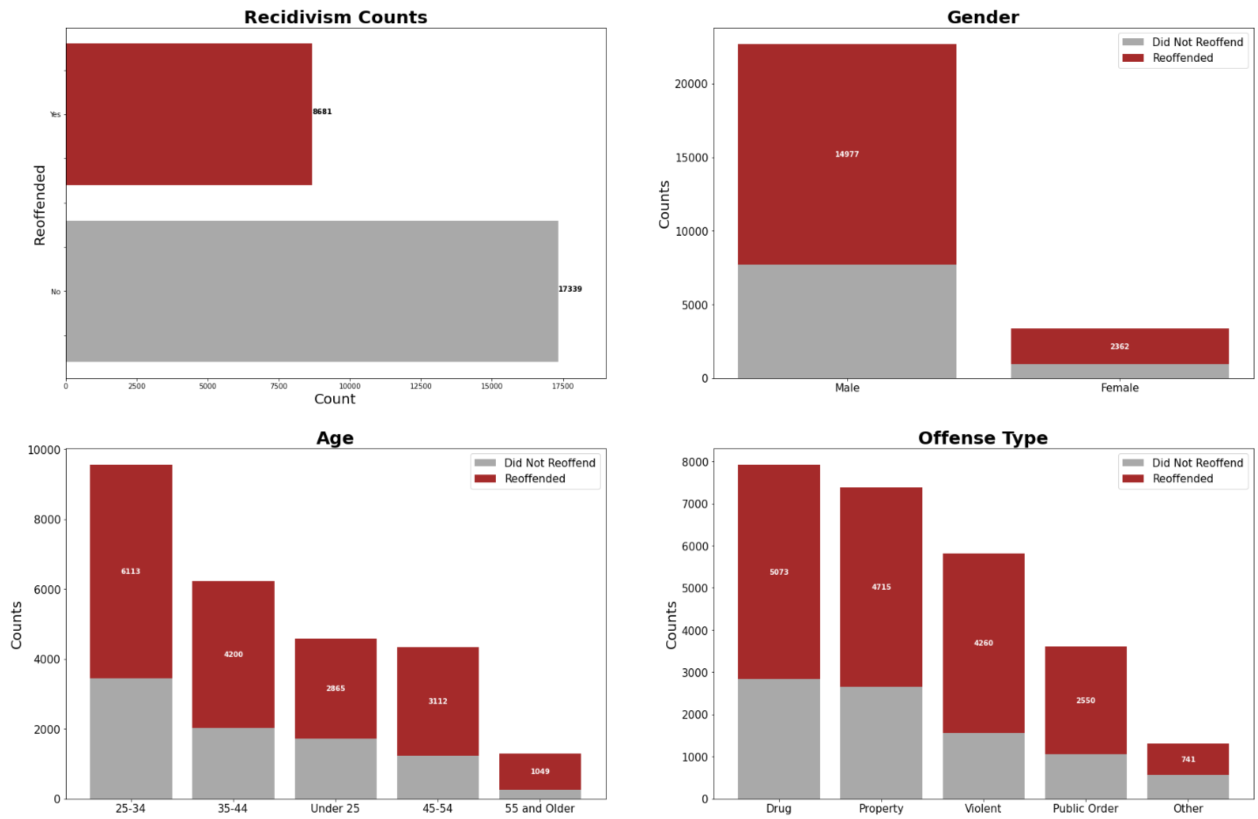


Figure 3: Preliminary Visualizations

<u>Data Transformation:</u>

Once all data cleaning was complete the data was transformed using one-hot encoding and then divided into a 70/30 train-test split. The original dataset had an imbalanced number of target variable records containing 66.6% non-reoffenders and only 33.4% reoffenders. This is inconsistent with the reported national average of 67.8% reoffenders. The reason for this inconsistency in unknown. Similar percentages of imbalance were identified in the training dataset and to prepare for data modeling, upsampling of the training target variable reoffender class was conducted.

Feature Selection

Feature selection is extremely important in machine learning primarily as it serves as a filtered technique to direct the use of variables to what is most efficient and effective with deference towards potential bias. For this project we elected to use a RandomForestClassifier as well as the SelectFromModel class from Scikit-learn's API. The RandomForestClassifier creates a set of decision trees from a randomly selected subset of training data and then aggregates the votes from different decision trees to decide the final class of the test object. The SelectFromModel class was used as it extracts the best features of a given dataset according to the importance of weights. These final extractions were then added to a final data frame that was used throughout the modeling. The one topic that needed to be managed was the potential bias in the data. When dealing with a sensitive issue such as this, the ethics of using the model to make decisions was something we had to address. The selection of features that we thought would help the model, excluding those that hold implicant bias overall, was a large lens that needed to be applied and maintained throughout this process. With the appropriate features selected, and the newly formed data frame ready to go, we moved on to model selection and evaluation.

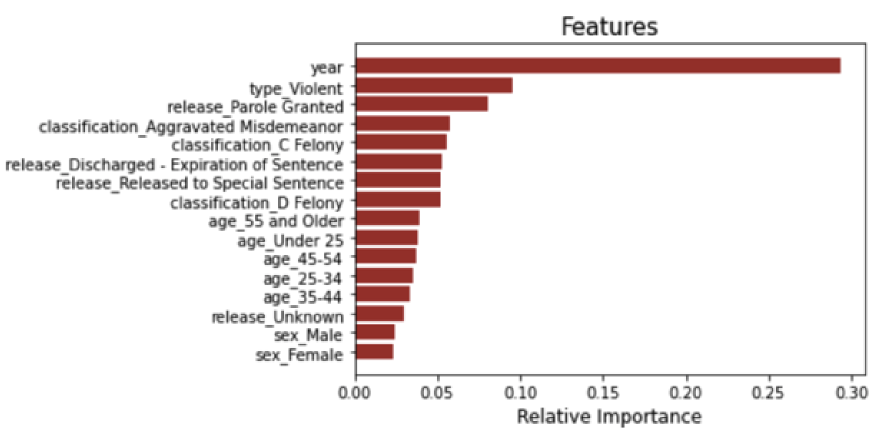


Figure 4: Feature Importance

Model Selection and Evaluation

Our model selection took some time to narrow down on exactly which methodology would result in the best overall product. With that said, trial and error played a large role throughout the process; analyzing the mechanics of each individual model, and then comparing results hoping to make incremental progress each time. For this project we identified multiple candidates: Logistic Regression, Random Forest, Naïve Bayes, and an Adaboost model to answer our topic problem.

With the independent variable being categorical in nature, we felt that a logistic regression model does well with this type of scenario as we were ultimately seeking a binary response. Logistic models don't require a linear relationship between dependent and independent variables, so this made this type of model a suitable candidate for the final product.

Originally selected to compute the prediction for the return to prison decision. The Random Forest model is suited for categorical variables, and the inclusion of a variable importance factor was a great addition to the overall analysis. With the variables appearing to not have an immediate linear relationship, this kind of model is also very tolerant in handling outliers and other unknowns that may be present also making this potentially an excellent option.

Our original assumption was that the features did not hold a particularly strong relationship. With that said, a Naïve Bayes model performs better in these circumstances than other models.

With Bayes theorem helping to find probability of a hypothesis given prior knowledge, it makes sense that this could have been an excellent choice for modeling.

Being an ensemble learning method that was originally designed to increase the efficiency of binary classifiers, an Adaboost model made an excellent choice for this project. Adaptive boosting serves to combine multiple weak classifiers in order to build a single strong one. Therefore, while it didn't appear that we had any specific features that acted as a strong classifier, Adaboost being layered into the existing ones served as a viable solution to this problem.

We evaluated each of the models by utilizing a 70/30 training and testing set. In our initial evaluations, we were sure to apply default parameters to establish a baseline metric in which we then could compare to. Hyperparameter tuning was then accomplished utilizing RandomizedSearchCV and the Adaboost model tuning was done using GridSearchCV.

Results

The AdaBoost model results reveal an accuracy rate of slightly better than pure chance, 66% when using the Test dataset. The metrics from the Training dataset produce the same or similar results. Additionally, no other model had an Accuracy score greater than 66%. Accuracy rate is only one of several metrics the team uses to evaluate the AdaBoost model's performance. The resultant Classification Report, ROC Curve, and Precision/Recall Curve, and Confusion Matrix provide additional metrics.

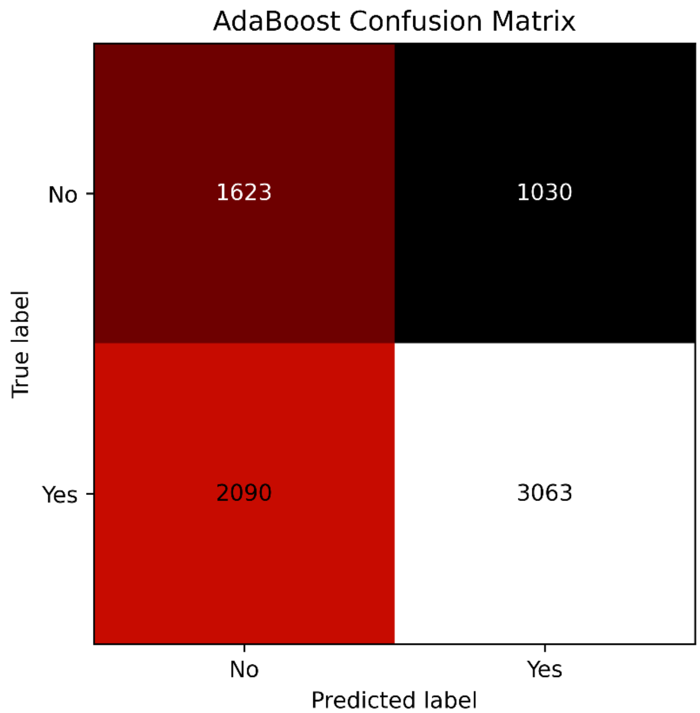


Figure 5: AdaBoost Confusion Matrix

Referring to figure Confusion Matrix generated using the *plot confusion matrix* module (https://scikit-learn.org, n.d.), the model correctly identified 1623 True Positive and 3063 True Negative predictions. Conversely, the model mis-identified 2090 False Positive and 1030 False Negative predictions.

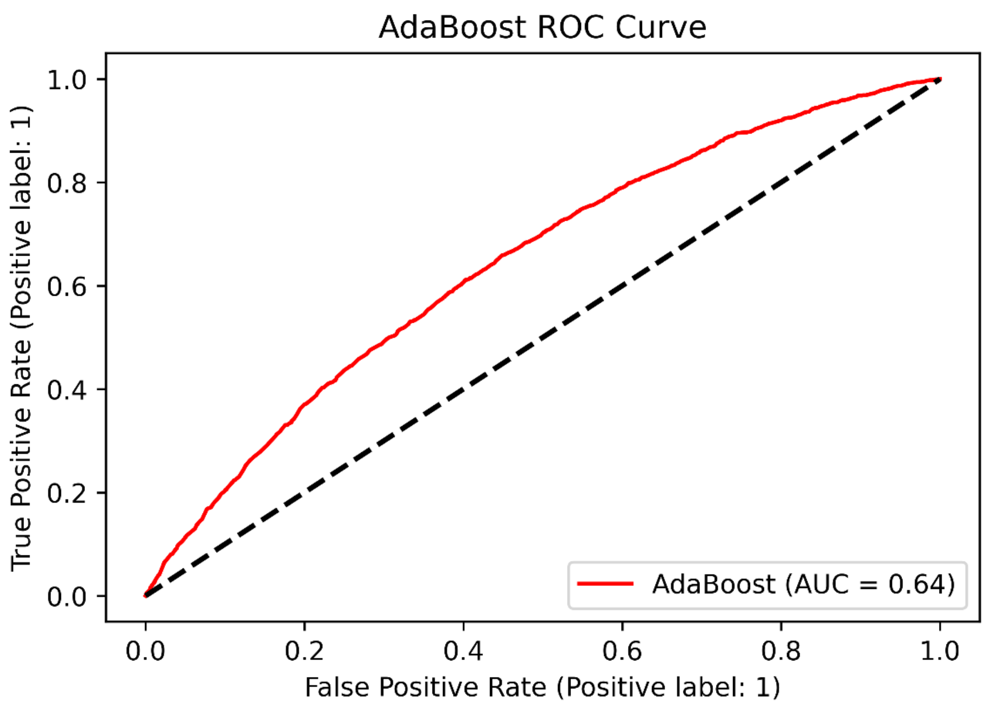
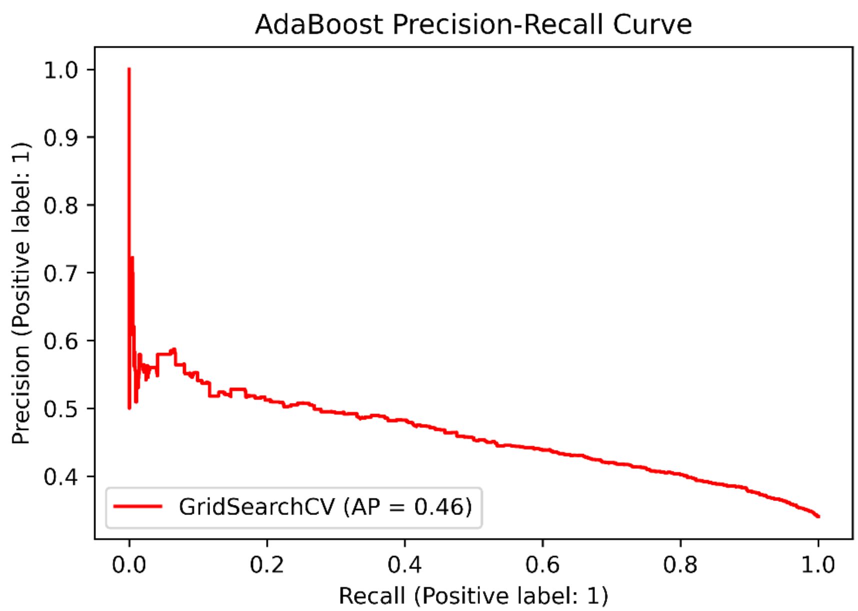


Figure 6: AdaBoost ROC Curve

Examining the ROC Curve, the ROC Curve reflects how the number of correctly classified positive results varies with the number of incorrectly classified negative results (Davis & Goadrich, 2021). An optimum ROC curve is close to the upper-left-hand corner. In contrast, our model's ROC Curve, figure AdaBoost ROC Curve, is close to the zero line and reflects the poor Accuracy Rate. Precision is a measure of the percentage of relevant results. Whereas Recall refers to the percentage of total relevant (Precision) results correctly classified by the model. An optimum Precision/Recall has the curve in the upper-right-hand corner (Davis & Goadrich, 2021). Referring to figure Precision/Recall Curve we see our model's curve is far from being in the upper right-hand-corner. The Micro Average Precision score over both classes is 0.46. From figure Classification Report, it shows a Weighted Average Precision score of 0.64 and a Weighted Average Recall score of 0.60. Looking at the F1-score, the results show a 0.66

for the Negative Class and a dismal 0.51 for the Positive class.   From the results, we can deduce that our model only identifies relevant values 64% of the time. Out of that 64%, the model is only able to classify correctly 60% of the predictions.



	precision	recall	f1-score	support
0	0.75	0.59	0.66	5153
1	0.44	0.61	0.51	2653
accuracy			0.60	7806
macro avg	0.59	0.60	0.59	7806
weighted avg	0.64	0.60	0.61	7806

Figure 7: AdaBoost Precision-Recall Curve and Details

Conclusion/Discussion

Of all the model algorithms used, none produced better metrics than the AdaBoost model. In fact, the team developed three additional models to better the results. The team developed a Stacked Ensemble model using a Logistics Regression, Naïve Bayes, and Random Forest algorithms. The Accuracy results were approximately 63%. The next model we developed was the Weighted Average Ensemble using Random Forest, Naïve Bayes, and AdaBoost algorithms. The Accuracy rate for this model is 66%. The last model is a Feed Forward Neural Network model using 3 layers and compiled with the Adam as the Optimizer and the Binary Cross Entropy as the loss function. Using these parameters produced an accuracy metric of 66%.

With such a low Accuracy and Precision Recall metrics, the AdaBoost model is not a suitable solution to predict Recidivism potential. A model used for a critical decision such as Recidivism needs to have an Accuracy Rate in the high 90 percentile range.

What are some possible reasons for the low accuracy and metrics? One reason may be the lack of or low correlation coefficients between variables. The variables did not provide statistically significant data relationships or patterns. Another possibility and is that the dataset does not include other significant variables that may provide the needed patterns and relationships for the prediction models.

Citing (Florida Department of Corrections, 2020) this report identifies several other variables that are possible candidates for inclusion into a Recidivism Model. Some of these variables include Inmate Incarceration Information, Inmate Educational Information, Inmate Criminal

History, and Inmate Social and Cognitive Factors. Within each major category are several specific attributes that may increase the possibility of producing a model with much improved accuracy scores. Examples of these attributes include Inmate homeless at Release, Inmate Employed at Release, Diagnosed with Mental Illness, Substance Abuse Severity Score, Critical Thinking Score, Criminal Associates Score, High School Certificate, Gang Member, and Inmate Participates in Faith based programs. In contrast to the IOWA Department of Corrections dataset which contains attributes pertaining to criminal activity.

Future Work

Predicting Recidivism is an interesting project and with the right dataset, creating a model that provides a high prediction score is attainable. To continue with this project, a dataset from the Florida Department of Corrections or one from another agency with similar attributes is a necessity. Additionally, using Neural Networks with tuned hyperparameters is one of the techniques to explore. For comparisons against the Neural Network model, we could use the same Ensemble methods or explore other algorithms to include in the Ensemble. For example, adding a XGBoost algorithm versus the Naïve Bayes is an option. Regardless of the model selections, one point this project has shown is the value of a viable dataset. This being a dataset that contains adequate variables for a machine learning model to recognize patterns and intrinsic data relationships.

Acknowledgements

We would like to thank Professor Brett Werner and all other professors that are part of Bellevue University's data science program. Without their direction, support, and guidance, this paper would not have been possible. We would like to thank Eric Seigel for inspiring the topic of our paper. Lastly, we would like to thank our families for sacrificing their time with us to allow us to pursue this degree for the betterment of ourselves.

References

Albon, C. (2018). Machine Learning with Python Cookbook: Practical Solutions From Preprocessing to Deep Learning. First Edition. O'Reilly Media, Inc.

Davis, J., & Goadrich, M. (2021). The Relationship Between Precision-Recall and ROC Curves. https://www.biostat.wisc.edu/~page/rocpr.pdf

Durose, M., Cooper, A., Snyder, H. (2014, April 22). Recidivism of Prisoners released in 30 states in 2005: Patterns From 2005 – 2010 – Update. Bureau of Justice Statistics. Retrieved from https://www.bjs.gov/index.cfm?ty=pbdetail&iid=4986.

Florida Department of Corrections. (2020, June). Florida Prison Recidivism Report:Releases from 2008 to 2018. http://www.dc.state.fl.us/pub/recidivism/2019-2020/FDC_AR2019-20.pdf

https://scikit-learn.org. (n.d.). Confusion matrix — scikit-learn 0.24.2 documentation. Retrieved May 31, 2021, from https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html#sphx-glr-auto-examples-model-selection-plot-confusion-matrix-py

Iowa Department of Correction. (2020, November 16). 3-Year Recidivism for Offenders Released from Prison in Iowa. Iowa Data. Retrieved from https://data.iowa.gov/Correctional-System/3-Year-Recidivism-for-Offenders-Released-from-Pris/mw8r-vqy4.