PREDICTIVE ANALYSIS OF LONG-TERM RISK FACTORS OF INCARCERATION FOR AT-RISK YOUTH

by
Kathleen Mason

A capstone submitted to the Johns Hopkins University in conformity with the requirement for the degree of Master of Science in Government Analytics

Baltimore, Maryland
August 2017

Abstract

The use of predictive analytics to prevent crimes, or predictive policing, is increasingly used by governments as part of criminal justice and law enforcement strategies today. Broadening the application of predictive analytics to encompass primary prevention strategies better informs government policy by leveraging the additional dimension of identifying risk factors that predispose youth to incarceration. To predict incarceration, a decision tree model with 72% accuracy was developed using data from the Prevention Program, a longitudinal intervention conducted on 900 first grade students in urban Baltimore Public Schools in 1993. A K-means clustering model was applied to the population to identify three archetype profiles: females, adapted males, and maladapted males. Overall classroom behavior, authority acceptance, and overall behavioral problem contributed most to the clusters. Results show that peer-rated and teacher-rated behavioral factors primarily influenced both predictive models. This indicates that peers and teachers are a valuable resource for identifying at-risk youth.

Table of Contents

1. Introduction

A combination of high crime rates and shrinking budgets have lead many large cities, such as Los Angeles, Baltimore, and Chicago, to explore predictive analytics methods as a possible solution to crime reduction. Governments combat crime using a combination of primary prevention, criminal justice, and law enforcement strategies. Primary prevention strategies are community interventions designed to increase solidarity and harmony within communities through public health initiatives. Predictive policing, or the application of predictive analytics to prevent and detect crime has gained the most traction as part of criminal justice and law enforcement strategies.[1] In contrast to criminal justice and law enforcement strategies, primary prevention programs have not garnered nearly the same interest in predictive analysis applications for three main reasons.

First, publicly available large datasets on longitudinal interventions for at-risk youth are limited, which consequently limits the ability of predictive policing applications in this domain. Second, historically, interventions generally fell within the realm of public health. Only in the past three decades have police broadened their approach to encompass more community-oriented and problem-solving strategies.[22,3] Lastly, primary prevention strategies traditionally have a more qualitative approach due

---

[1] Bachner, Jennifer. Predictive policing: Preventing crime with data and analytics. IBM Center for the Business of Government, 2013.
[2] Greene, Jack R. "Community policing in America: Changing the nature, structure, and function of the police." *Criminal justice* 3, no. 3 (2000): 299-370.

Reisig, Michael D., and Robert J. Kane, eds. *The Oxford handbook of police and policing*. Oxford University Press, 2014.

to their subjective nature whereas predictive analytics heavily relies on quantitative data.

Through public funding, though, quantitative data on interventions such as the Johns Hopkins Prevention Program, have been collected for longitudinal studies sponsored by government organizations, including the U.S. Center for Disease Control and Prevention (CDC) and National Institutes of Health (NIH).[3] The Prevention Program is a twenty-year longitudinal study for an intervention with at-risk youth in inner city Baltimore public schools conducted by the Johns Hopkins School of Public Health. This dataset collected a myriad of variables on participants from first grade until around age 26. Primary prevention strategies are at the intersection of public health and public safety, and may benefit from insight into long-term risk factors for incarceration.

2. Literarature Review

2.1 Definition of Predictive Policing

Predictive policing is defined as the use of predictive analytics on large datasets to prevent crime.[4] Predictive analysis employs advanced statistical methods and data mining techniques to estimate unknown variables. The four main objectives of predictive policing analysis are predicting crimes, predicting offenders, predicting perpetrators' identities, and predicting victims of crime. Predictive policing has shown

---

[3] Johns Hopkins Bloomberg School of Public Health. "Center for Prevention and Early Intervention Home Page." Johns Hopkins Bloomberg School of Public Health. August 23, 2010. Accessed June 06, 2017. http://www.jhsph.edu/research/centers-and-institutes/johns-hopkins-center-for-prevention-and-early-intervention/index.html.
[4] Bachner, 8.

Perry, Walt L. Predictive policing: The role of crime forecasting in law enforcement operations. Rand Corporation, 2013.

Berk, Richard. "Algorithmic criminology." *Security Informatics* 2, no. 1 (2013): 5.

promise to be an effective tool in fighting crime. This effectiveness is predicated on the type of crime, however. The systematic, or predictable component, of crime is where the highest impact can seen.[8] Burglaries, for example, are highly systematic and rarely random. It is not surprising, for example, that the Santa Cruz police department reported a 27% reduction in burglary rates six months after predictive policing methods were first implemented and the Alhambra, CA police department reported a 32% reduction in robberies nine months after predictive policing methods were implemented.[5] Non-systematic crime, however, such as spontaneous murder due to an emotional outrage, would not benefit from predictive policing. Overall, police departments across the U.S. are reporting significant reductions after implementing predictive policing strategies; however, the extent to which the impact of these strategies can be causally linked to predictive policing is difficult to determine.[6]

2.2 Primary Prevention Strategies

Primary prevention strategies were initially introduced in the early 1980s as part of community policing reform efforts to improve police-community relations and as a broader effort to find new strategies to reduce crime.[7] These prevention strategies gained traction and community policing became well-integrated into crime prevention strategies by the 1990s. Prevention strategies police departments have adopted are situational crime

[5] "Proven Results of our Predictive Policing Software." PredPol. Accessed June 06, 2017. http://www.predpol.com/results/.
[6] Kirkpatrick, Keith. "It's not the algorithm, it's the data." *Communications of the ACM* 60, no. 2 (2017): 21-23.
[7] Reisig & Kane, 148.

prevention, crime prevention through environmental design, community crime prevention, business crime prevention, and youth-oriented prevention.[8] With situational crime prevention, preventive measures are adapted according to the circumstances of the situation. Crime prevention through environmental design seeks to minimize physical characteristics in the environment that inherently invite crime through the design of communities. Community crime prevention is a cooperation between police departments and residents with the common goal of preventing crime, such as neighborhood watch programs. Business crime prevention is the cooperation between police departments and local businesses, where police provide information and guidance to better secure their businesses against crime. Under the umbrella of community policing, youth-oriented prevention efforts are a partnership between police and community members to develop programs to prevent juvenile crime and reduce recidivism.

Youth-oriented crime prevention efforts have evolved to become a fundamental part of the justice system in the United States. The National Institute of Justice maintains a library of research on 239 intervention programs for preventing juvenile delinquency as part of its commitment to public safety.[9] One such program, the Good Behavior Game, was developed out of the Johns Hopkins Prevention Program study.[10] Similarly, the National Center for Injury Prevention and Control (NCIPC) was established in 1992 by the CDC to combat violence. The Division of Violence

---

[8] Cordner, Gary. "Community policing: Principles and elements." *National Institute of Justice, Washington, DC* (1996).
[9] "CrimeSolutions.gov." Topic: Delinquency Prevention - CrimeSolutions.gov. Accessed June 06, 2017. https://www.crimesolutions.gov/TopicDetails.aspx?ID=62.
[10] Kellam, Sheppard G., George W. Rebok, Nicholas Ialongo, and Lawrence S. Mayer. "The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically‐based preventive trial." *Journal of Child Psychology and Psychiatry* 35, no. 2 (1994):259-281.

Prevention (DVP), one of three divisions within the NCIPC, also leads the Task Force for Community Preventive Services whose primary goal is to conduct, identify, and evaluate effective prevention strategies for youth violence. While both of these programs are at the federal level, state and local level law enforcement across the nation have embraced community policing as a valuable component to preventing crime.

2.3 Long-term Risk Factors

Long-term risk factors are factors in childhood that predispose an individual to incarceration later in life. Although a multitude of long-term risk factors, such as academic absenteeism, academic failure, drug addiction, gang membership, and sociodemographic characteristics, have been linked to delinquency, poverty is especially pernicious to youth outcomes.[11] According to the National Center for Children in Poverty, about 15 million or 21% of children in the United States lived in poverty in 2015, yet comprise over one-third of all individuals living in poverty in the United States.[12] School truancy, academic failure, and drug addiction are heightened risk factors

---

[11] Teasley, Martell L. "Absenteeism and truancy: Risk, protection, and best practice implications for school social workers." *Children & Schools* 26, no. 2 (2004): 117-128.

Alexander, Karl, Doris Entwisle, and Nader S. Kabbani. "The Dropout Process in Life Course Perspective." *Teachers College Record* 103 (2001): 760-882.

Sampson, Robert J., and John H. Laub. "A life-course view of the development of crime." *The Annals of the American Academy of Political and Social Science* 602, no. 1 (2005): 12-45.

Hill, Karl G., James C. Howell, J. David Hawkins, and Sara R. Battin-Pearson. "Childhood risk factors for adolescent gang membership: Results from the Seattle Social Development Project." *Journal of Research in Crime and Delinquency* 36, no. 3 (1999): 300-322.

[12] Jiang, Yang, Maribel R. Granja, and Heather Koball. "Basic Facts about Low-Income Children, Children under 18 Years, 2015." (2017).

for children raised in poverty with familial instability.[13] This increased risk directly translates into higher school dropout rates, and, consequently, higher rates of juvenile delinquency.

Increased levels of education have been strongly correlated with lower rates of incarceration and also partially accounts for the racial disparity in incarcerated black males.[14] Moreover, educational attainment levels, epecially post-secondary education, of incarcerated individuals is much lower compared with the general public.[15] One study estimates that among black men born between 1965 and 1969, 30% of those who graduated with a high school degree and almost 60% of those that did not were incarcerated at least once by 1999.[16] The disadvantages conferred to formerly incarcerated individuals are many and reintegration with family, community, and society is challenging, leading to high recidivism rates.

Another precursor to crime is age; young offenders under the age of 13 are much more likely to become violent and chronic juvenile offenders compared with adolescent offenders.[17] However, if early interventions are successful at reducing maladaptive

---

[13] Arditti, Joyce A., Jennifer Lambert−Shute, and Karen Joest. "Saturday morning at the jail: Implications of incarceration for families and children." *Family relations* 52, no. 3 (2003): 195-204.
[14] Lochner, Lance, and Enrico Moretti. "The effect of education on crime: Evidence from prison inmates, arrests, and self-reports." *The American Economic Review* 94, no. 1 (2004): 155-189.
[15] Brazzell, Diana, Anna Crayton, Debbie A. Mukamal, Amy L. Solomon, and Nicole Lindahl. "From the classroom to the community: exploring the role of education during incarceration and reentry." *Urban Institute (NJ1)* (2009).
[16] Western, Bruce, and Becky Pettit. "Black-white wage inequality, employment rates, and incarceration 1." *American Journal of Sociology* 111, no. 2 (2005): 553-578.
[17] Burns, Barbara J., James C. Howell, Janet K. Wiig, Leena K. Augimeri, Brendan C. Welsh, Rolf Loeber, and David Petechuk. "Treatment, services, and intervention programs for child delinquents. Child Delinquency Bulletin Series." (2003).

behaviors in the early, formative years, young offenders are less likely to exhibit these behaviors. Furthermore, research has demonstrated that imposing a minimum drop-out age of 18 is associated with lower arrest rates for 16 to 18 year-olds.[18]

2.4 Predictive Analysis

Current predictive applications to law enforcement and criminal justice seek to prevent crime with short-term risk factor analysis and reduce recidivism with post-incarceration risk factors, respectively. Broadening the application of predictive analytics to encompass primary prevention strategies may provide insight to government policy by leveraging the additional dimension of long-term risk factors. Given that large datasets on longitudinal interventions are becoming available as well as recent evidence of successful intervention programs, valuable insights could be gained from machine learning and predictive analytics applications.

This analysis seeks to extend the application of predictive analytics to primary prevention strategies. Two types of predictive analytics, decision-tree modeling and clustering modeling will be used to not only identify predictive factors, but to also identify archetype profiles of individuals at risk for incarceration. Using the Prevention Program intervention dataset, this predictive analysis examines sociodemographic, behavioral, environmental, and policy factors as potential predictors of incarceration and drug use with a supervised decision-tree analysis. Because previous analyses have already investigated the effect of the intervention, this variable will be excluded from the

[18] Anderson, D. Mark. "In school and out of trouble? The minimum dropout age and juvenile crime." *Review of Economics and Statistics* 96, no. 2 (2014): 318-331.

analysis.[19] The unsupervised clustering analysis will be used as a method to predict offenders as candidates for intervention and primary prevention strategies. The results of this analysis can be used to strengthen community policing programs as well as improve public health intervention programs for at-risk youth.

3. Data and Methods

The Prevention Program experimental panel dataset ranging from 1993-2013 will be used to analyze the relationship between incarceration and factors that predispose youth to crime. The dataset contains incarcerations and predisposition factors collected over twenty years from 1993 to 2013 as part of the Prevention Program is a longitudinal study conducted on 900 students in inner city Baltimore Public Schools. The outcomes of individuals following a classroom-based intervention of first graders with compiled annual interviews on the participants, teachers, and parents will be examined to identify predictive factors. This dataset is being provided by the Center for Prevention and Early Intervention of the Johns Hopkins Bloomberg School of Public Health.

3.1 Independent Variables

[19] Ialongo, Nick, Jeanne Poduska, Lisa Werthamer, and Sheppard Kellam. "The distal impact of two first-grade preventive interventions on conduct problems and disorder in early adolescence." *Journal of Emotional and behavioral disorders* 9, no. 3 (2001): 146-160.

Ialongo, Nicholas S., Lisa Werthamer, Sheppard G. Kellam, C. Hendricks Brown, Songbai Wang, and Yuhua Lin. "Proximal impact of two first−grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior." *American journal of community psychology* 27, no. 5 (1999): 599-641.

Dolan, Lawrence J., Sheppard G. Kellam, C. Hendricks Brown, Lisa Werthamer-Larsson, George W. Rebok, Lawrence S. Mayer, Jolene Laudolff, Jaylan S. Turkkan, Carla Ford, and Leonard Wheeler. "The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement." *Journal of Applied Developmental Psychology* 14, no. 3 (1993): 317-345.

For this study, the analysis focused on fifty-seven variables related to criminal disposition covering socioeconomic factors, school records, psychological well-being, social adaptation status, and mediators and moderators of behavior obtained when children were in first grade. Socioeconomic factors included information such as race and gender. School records include attendance and Comprehensive Tests of Basic Skills scores for math and reading comprehension. Psychological well-being estimators included emotions, depression, and anxiety. Social adaptation status includes surveys on peer, parent, and teacher assessments of participant conduct, participant attention and concentration, and participant likability and shyness. To obtain peer-ratings, the Peer Assessment Inventory provided feedback based on aggressive behavior, shyness, and likability. To obtain teacher ratings, the Teacher Observation of Child Adaptation-Revised (TOCA-R) was used to assess a student performance on the sub-scales of authority acceptance, attention and concentration problems, shyness, hyperactivity, peer likability, and impulsivity. Total problem score is the mean of all sub-scales of the TOCA-R. Overall classroom behavior is a global teacher-rating based on participant behavior in the past three weeks. Mediating and moderating factors include parent involvement and discipline as well as family mental health factors, such as mental illness in the family or drug use of a family member.

3.2 Data Pre-Processing

After evaluating all variables, feature selection reduced the number of variables for analysis to nineteen factors.[20] The exclusion criteria used to evaluate features were

[20] See Appendix A - Variables Selected for Analysis

unbalanced fields and low variation coefficient variation. In all analyses, missing values were imputed. Additionally, twenty-six anomalous records were discovered based on an anomaly index of two.[21] A binary anomaly field was derived based on whether a record was classified as an anomaly to see if anomalies have an impact on predicted incarceration.

3.3 Target Variable

This analysis will examine the impact these long-term factors have on incarceration. Incarceration is evaluated as a binary variable as incarcerated or never incarcerated. The distribution of incarceration was unbalanced, with 90% not incarcerated and 10% incarcerated. To optimize the model, this field was boosted, resulting a more balanced field with 59% not incarcerated and 41% incarcerated. Records with missing values in this target variable were omitted.

4. Results

4.1 Decision Tree Incarceration Prediction Model

Using a 70% train and 30% test split, the Classification and Regression Tree (C&RT) standard algorithm produced a decision tree model with 86% accuracy on the training set and 72% accuracy on the testing set resulting in overfitting of the model by 14%.

[21] An anomaly index is the ratio of the group deviation index to its average over the cluster that the case belongs to.

The confusion matrix for the test set, shown in Figure 1, shows that the number of participants expected to be incarcerated were 56, but only 12 were actually incarcerated. While this is an overestimation of those incarcerated, this is still preferable to equip those participants most at-risk for incarceration than underestimate those who may need the additional support.

Boosting the model resulted in higher accuracy at 82% compared with the original model which had an accuracy of 72%. Although boosting the model improved accuracy, it also increased overfitting to the data to 17%.[22] Regardless, the model offers sufficient accuracy for prediction using the decision tree.

		Prediction Class	
		Yes	No
	Yes	12	12
	No	44	135

Figure 1. Confusion Matrix, Test Set

[22] See Appendix B - Technical Explanation.

4.2 Features of Incarceration Prediction Model

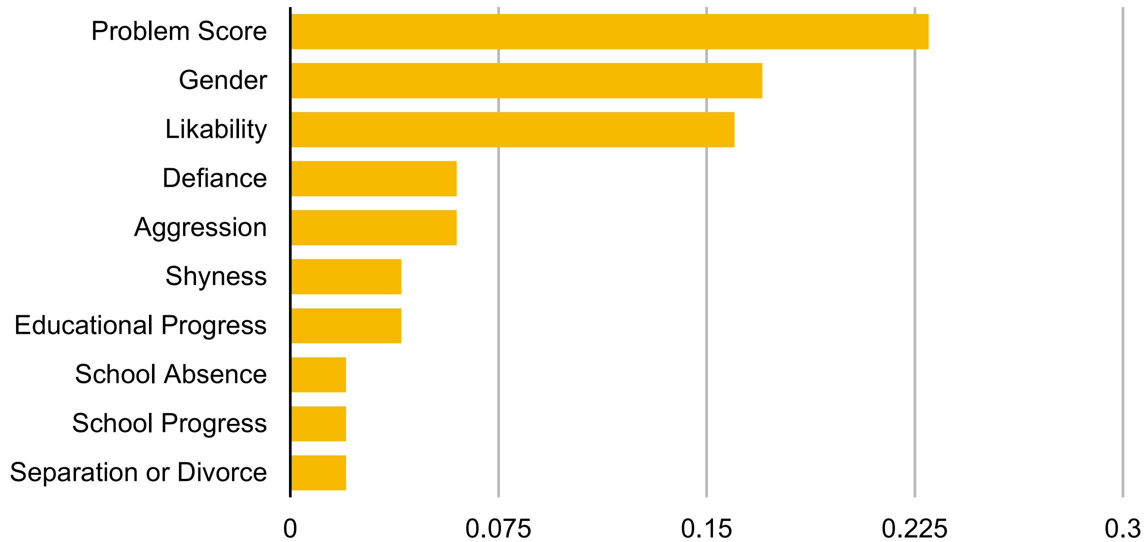


Figure 2. Predictor variables of the decision tree model ranked by importance.

The top three contributing features of the decision tree model were total problem score, gender, and peer-rated likability. With an importance score of 0.23, total problem score far exceeded both gender with an importance of 0.17 and peer social likability preference score with an importance of 0.16. That said, the importance of gender aligns with the data, as 88% of incarcerated individuals were male as shown Figure 3. The importance of these factors suggests that teachers and peers are a valuable resource for identifying at-risk youth.

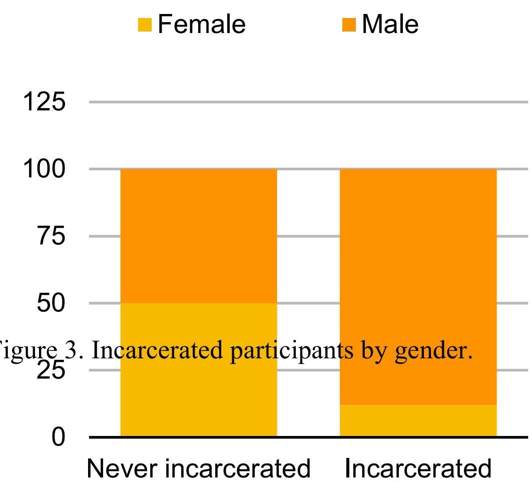
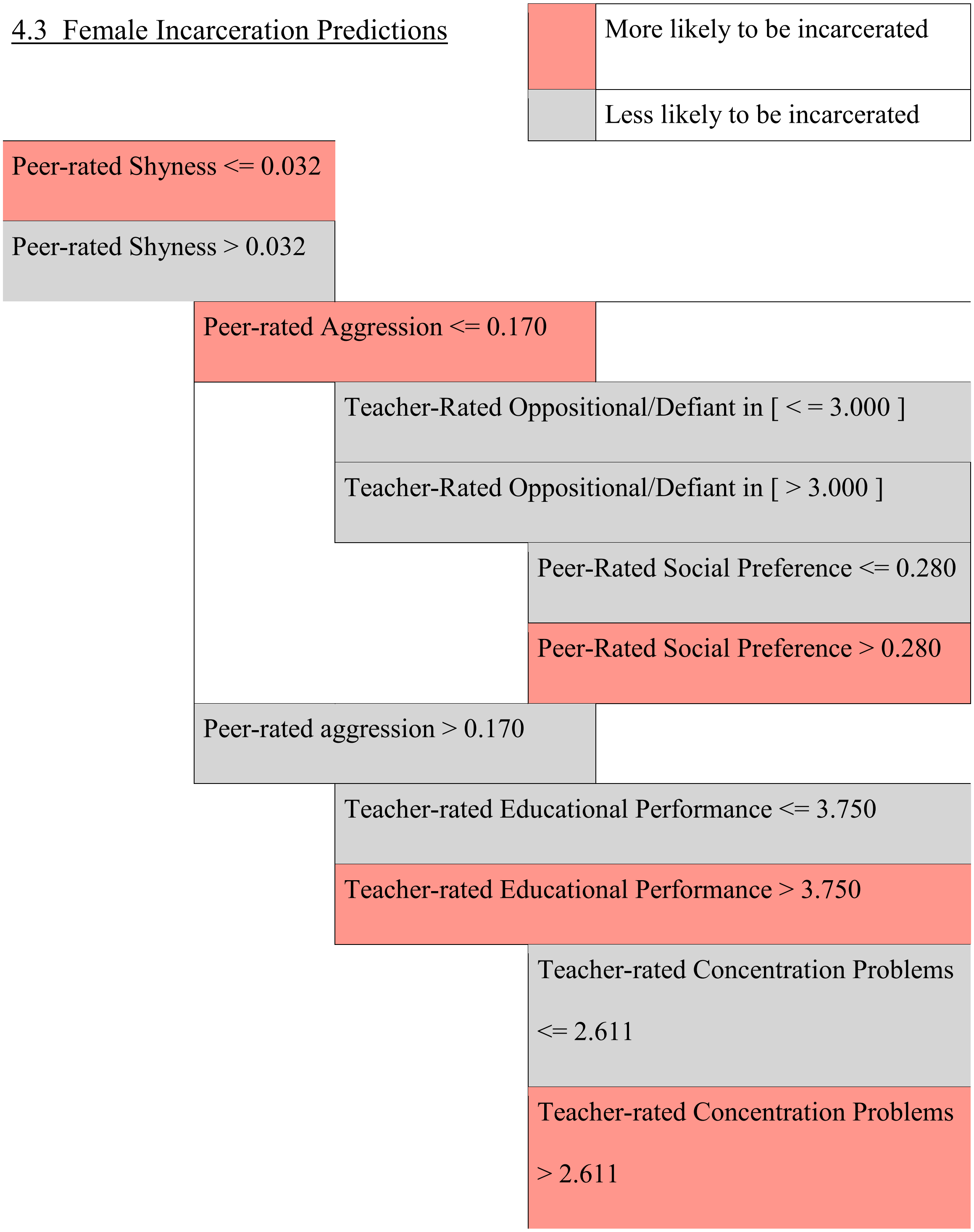


Figure 3. Incarcerated participants by gender.

Other factors that ranked within the top half were aggression, defiance, shyness, education progress, school progress, parent separation or divorce, and school absence.

Given the first split in the decision tree was by gender, the analysis will first evaluate features of female incarceration and then of male incarceration.

4.3 Female Incarceration Predictions



4.3 Female Incarceration Predictions

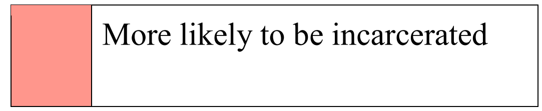
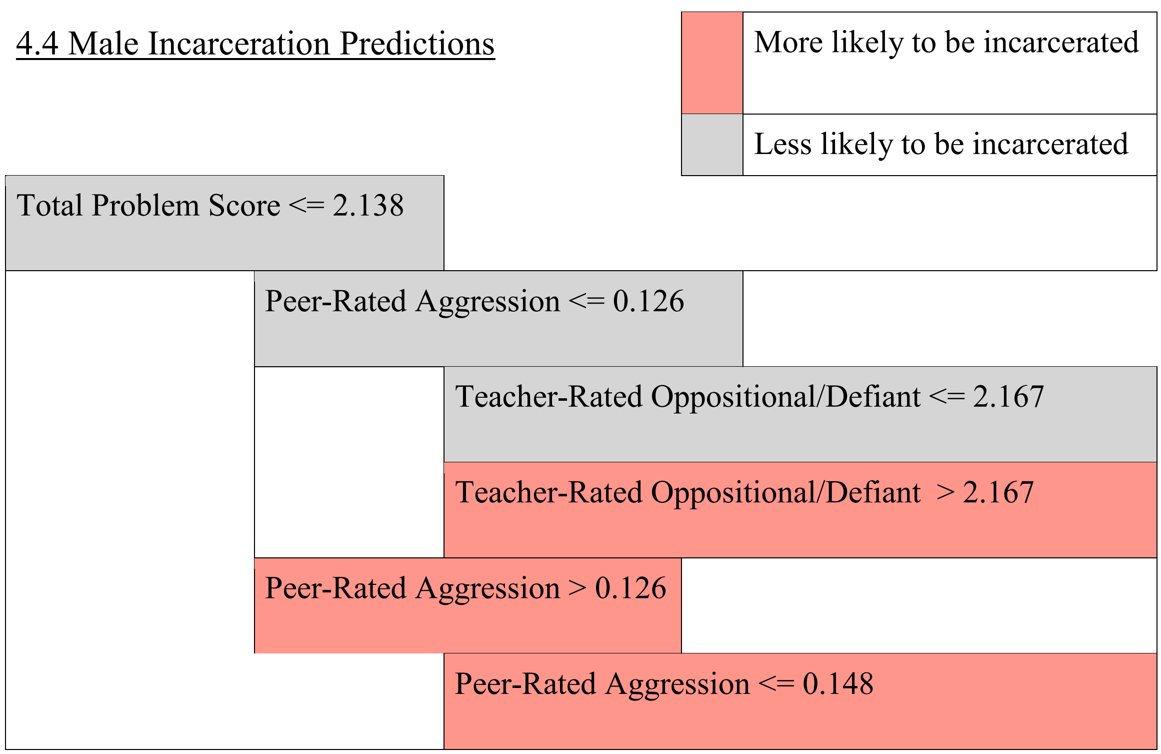


Figure 3. Female Incarceration Decision Tree Prediction Model

Female participants with a mean peer rating of shyness equal to or less than 0.032 were likely to become incarcerated. Of female participants with a mean percent peer rating of shyness greater than 0.032, those with a peer-rated aggression equal to or less than 0.17, teacher-rated oppositional defiant behavior greater than 3, and peer-rated social preference of likability greater than 0.280 were predicted to become incarcerated. Of female participants with peer-rated shyness less than 0.032 and peer-rated aggression less than 0.17, those with a teacher-rated educational performance greater than 3.75 and teacher-rated concentration problems over 2.611 were likely to become incarcerated.

4.4 Male Incarceration Predictions



4.4 Male Incarceration Predictions

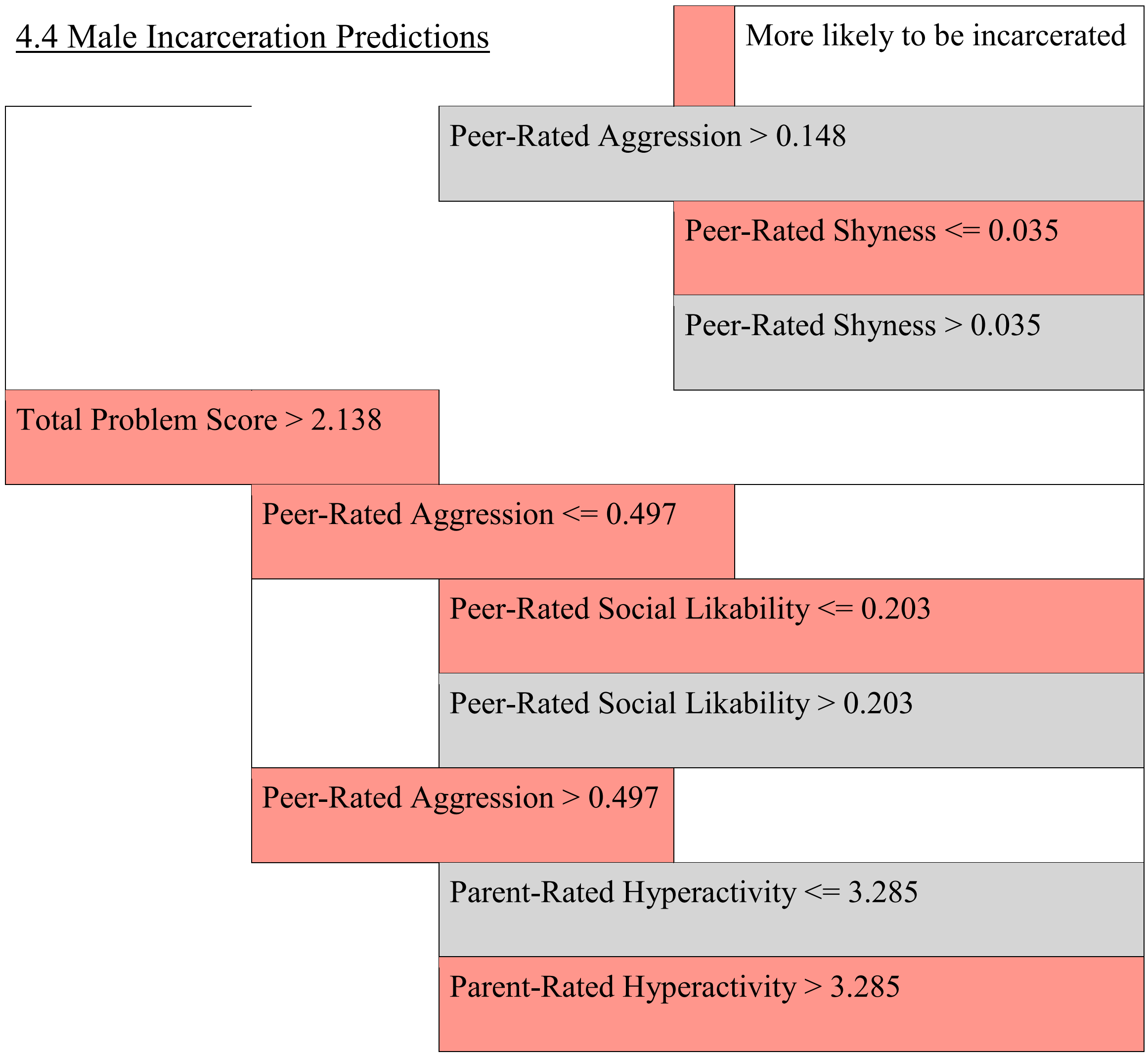


Figure 4. Male Incarceration Decision Tree Prediction Model

Of male participants with a total problem score less than 2.138, those with a peer-rated aggression score equal to or less than 0.126, and teacher-rated oppositional/defiant score greater than 2.167 were likely to be incarcerated. Those with a peer-rated aggression greater than 0.126 and peer-rated shyness equal to or less than 0.035 were predicted to be incarcerated.

Of male participants with a total problem score greater than 2.138, those with a peer-rated aggression score equal to or less than 0.497 and a peer-rated social likability score equal to or less than 0.203 were likely to become incarcerated. Those with a peer-rated aggression greater than 0.497 and parent-rated hyperactivity greater than 3.285 were also predicted to become incarcerated.

4.5 Archetype Profiles from Clustering

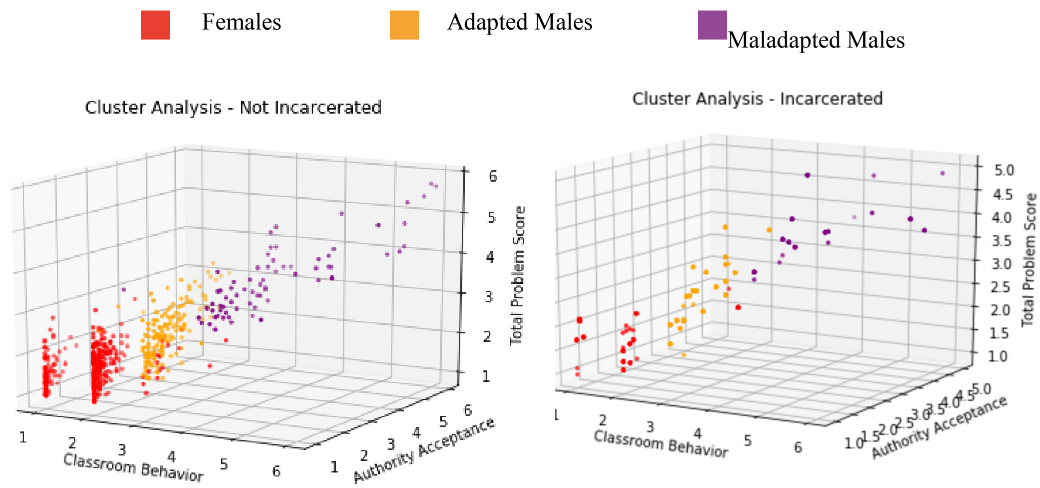
Using K-means clustering algorithm, three clusters were identified and their characteristics are ranked by predictor importance in Figure 5, Cluster Comparison. The initial K-means model revealed that the most important predictors were overall classroom behavior, total problem score, teacher-rated authority acceptance, and teacher-rated oppositional, defiant behavior. The factors that contributed least to the initial model and subsequently eliminated as inputs were school absence, peer-rated social preference, peer-rated shyness, parent-rated impulsivity, anomaly, and child experienced separation or divorce. Eliminating these inputs improved the silhouette of the model from poor cluster quality at 0.1 to fair cluster quality at 0.3. The first cluster represents females and contains 47% of data points. The high percentage of males being incarcerated at some point in their lives is highlighted by the second and third clusters, comprised primarily of males. The second cluster, containing 36% of data points, represents adapted males, and the third cluster, containing 17% of data points represents extremely maladapted males.

The first cluster is comprised females, characterized by good behavior, good school progress, high likability, and high education performance. This group was also

Figure 5. Cluster Comparison. Factors are ranked by

Figure 5. Cluster Comparison. Factors are ranked by predictor importance.

…entration, hyperactivity and conduct-disorder behavior. 23% participants in this cluster will be incarcerated at least once in their lives. While less than desirable, this is still significantly better than the incarceration



rates for the other two clusters, as can be seen in Figure 6.

Figure 6. Non-incarcerated and incarcerated participants by cluster based on top three predictors: Classroom behavior, authority acceptance, and total problem score.

The second cluster, comprised of adapted males with fair classroom behavior had higher total problem scores, higher impulsivity, aggression, and exhibited more oppositional, defiant behavior. Participants in this cluster also received lower scores in both social preference and education performance. 56% of participants in this cluster were incarcerated at some point in their lives.

The third cluster comprised of maladapted males with poor behavior and poor school performance. Participants in this cluster received significantly outscored the other clusters in total problem scores, impulsivity, aggression, and oppositional, defiant behavior. Of the participants in this group, 58% of this group have been incarcerated at some point in their lives. Interestingly, the incarceration rate for this cluster is only slightly higher than that adapted males, indicating that incarceration risk is higher overall for males, regardless of cluster classification.

5. Conclusion

The results of the analysis indicate that teacher and peer evaluations are a valuable resource in identifying at-risk youth. Gender was a predominant influence on both the decision tree and clustering models. In the decision tree model, the initial split on gender variable allowed for the identification of risk factors associated with each sex. With the clustering model, the influence of gender was reflected in the composition of the resulting clusters. Other contributing variables with high predictor importance in both models were aggression, likability, defiance, and educational performance.

5.1 Decision Tree Model

The decision tree model revealed that females perceived by peers to exhibit little shyness are predisposed to incarceration. In contrast, females perceived to exhibit more shyness and defiant behavior with less aggression and low likability were predicted to become incarcerated. Additionally, females with more shyness, more aggression, higher educational performance and higher concentration problems were likely to be incarcerated.

For males, total problem score was the most significant feature, and is not unsurprising this factor played a large role in clustering model, given the gender distribution of the incarcerated participants. Males with higher total problem scores, perceived to exhibit high aggression, and high parent-rated hyperactivity were predisposed to incarceration. Of male participants with a higher total problem score, lower perceived aggression and low likability were also at-risk for becoming incarcerated. In comparison, male participants with a lower total problem score were less likely to become incarcerated. Of those participants with a low total problem score, those with low perceived aggression but high oppositional, defiant behavior were more likely to be incarcerated as well as those with higher peer-rated aggression and low peer-rated shyness.

5.2 Clustering Model

The clustering algorithm produced three clusters, well-adapted females, adapted males and maladapted males. The most important predictors were primarily teacher-rated factors and the top three were overall classroom behavior, total problem score, and authority acceptance. The females cluster exhibited good overall classroom behavior, had low total problem scores, and generally accepted authority. With fair classroom behavior scores, some difficulty with authority acceptance, and higher total problem scores, adapted males tend to fall between females and maladapted males. The difference in scores between adapted males and females clusters mimic gender norms. Maladapted males were characterized by much higher total problem scores, poor overall classroom behavior, and significant difficulty with authority acceptance. Even though two distinct clusters of male profiles were formed, both groups still had equally high incarceration rates, following the broader trend of high incarceration rates for black males in America.

5.3 Limitations and Future Research

There are a couple of limitations to consider with this analysis. First, this analysis was performed on inner city Baltimore public school pre-dominantly black students. Further research is needed to determine if key findings of this analysis are unique to this population. Second, although 72% is an acceptable level of prediction for a model, this accuracy rate could be improved. The overfitting to the model by 14% indicates some irrelevant data points are being taken into account and may yield lower accuracy on future empirical application. Lastly, the silhouette quality of the cluster analysis was leaves room for improvement. Although 0.3 is fair quality, a more optimal silhouette would range between 0.5 and 1.

Since this analysis focused on incarceration, one opportunity for future exploration is predictive factors of recidivism. In addition recidivism, the Prevention Program dataset contains many other sociodemographic, neighborhood, policy-related variables that may be further studied. Given that the prevention program dataset is a panel dataset, the predictive potential of factors can be evaluated for both short-term and long-term trends. Lastly, a meta-analysis using data from multiple urban neighborhoods may also serve as baseline for future studies.

References

Alexander, Karl, Doris Entwisle, and Nader S. Kabbani. "The Dropout Process in Life Course Perspective." Teachers College Record 103 (2001): 760-882.

Anderson, D. Mark. "In school and out of trouble? The minimum dropout age and juvenile crime." Review of Economics and Statistics 96, no. 2 (2014): 318-331.

Arditti, Joyce A., Jennifer Lambert- Shute, and Karen Joest. "Saturday morning at the jail: Implications of incarceration for families and children." Family relations 52, no. 3 (2003): 195-204.

Bachner, Jennifer. Predictive policing: Preventing crime with data and analytics. IBM Center for the Business of Government, 2013.

Berk, Richard. "Algorithmic criminology." Security Informatics 2, no. 1 (2013): 5. "Proven Results of our Predictive Policing Software." PredPol. Accessed June 06, 2017. http://www.predpol.com/results/.

Brazzell, Diana, Anna Crayton, Debbie A. Mukamal, Amy L. Solomon, and Nicole Lindahl. "From the Classroom to the Community: Exploring the Role of Education during Incarceration and Reentry." Urban Institute (NJ1) (2009).

Burns, Barbara J., James C. Howell, Janet K. Wiig, Leena K. Augimeri, Brendan C. Welsh, Rolf Loeber, and David Petechuk. "Treatment, Services, and Intervention Programs for Child Delinquents. Child Delinquency Bulletin Series." (2003).

Cordner, Gary. "Community policing: Principles and elements." National Institute of Justice, Washington, DC (1996).

"CrimeSolutions.gov." Topic: Delinquency Prevention - CrimeSolutions.gov. Accessed June 06, 2017. https://www.crimesolutions.gov/TopicDetails.aspx?ID=62.

Dolan, Lawrence J., Sheppard G. Kellam, C. Hendricks Brown, Lisa Werthamer-Larsson, George W. Rebok, Lawrence S. Mayer, Jolene Laudolff, Jaylan S. Turkkan, Carla Ford, and Leonard Wheeler. "The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement." Journal of Applied Developmental Psychology 14, no. 3 (1993): 317-345.

“Future Trends in Predictive Policing.” Police Executive Research Forum. 2014. Future Trends in Policing. Washington, D.C.: Office of Community Oriented Policing Services.

Greene, Jack R. "Community policing in America: Changing the nature, structure, and function of the police." Criminal justice 3, no. 3 (2000): 299-370.

Hill, Karl G., James C. Howell, J. David Hawkins, and Sara R. Battin-Pearson. "Childhood risk factors for adolescent gang membership: Results from the Seattle Social Development Project." Journal of Research in Crime and Delinquency 36, no. 3 (1999): 300-322.

Ialongo, Nick, Jeanne Poduska, Lisa Werthamer, and Sheppard Kellam. "The distal impact of two first-grade preventive interventions on conduct problems and disorder in early adolescence." Journal of Emotional and behavioral disorders 9, no. 3 (2001): 146-160.

Ialongo, Nicholas S., Lisa Werthamer, Sheppard G. Kellam, C. Hendricks Brown, Songbai Wang, and Yuhua Lin. "Proximal impact of two first- grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior." American journal of community psychology 27, no. 5 (1999): 599-641.

Jiang, Yang, Maribel R. Granja, and Heather Koball. "Basic Facts about Low-Income Children, Children under 18 Years, 2015." (2017).

Johns Hopkins Bloomberg School of Public Health. "Center for Prevention and Early Intervention Home Page." Johns Hopkins Bloomberg School of Public Health. August 23, 2010. Accessed June 06, 2017.http://www.jhsph.edu/research/centers-and-institutes/johns-hopkins-center-for-prevention-and-early-intervention/index.html.

Kellam, Sheppard G., George W. Rebok, Nicholas Ialongo, and Lawrence S. Mayer. "The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically- based preventive trial." Journal of Child Psychology and Psychiatry 35, no. 2 (1994): 259-281.

Kirkpatrick, Keith. "It's not the algorithm, it's the data." Communications of the ACM 60, no. 2 (2017): 21-23.

Lochner, Lance, and Enrico Moretti. "The effect of education on crime: Evidence from prison inmates, arrests, and self-reports." The American Economic Review 94, no. 1 (2004): 155-189.

Perry, Walt L. Predictive policing: The role of crime forecasting in law enforcement operations. Rand Corporation, 2013.

Reisig, Michael D., and Robert J. Kane, eds. The Oxford handbook of police and policing. Oxford University Press, 2014.

Sampson, Robert J., and John H. Laub. "A life-course view of the development of crime." The Annals of the American Academy of Political and Social Science 602, no. 1 (2005): 12-45.

Teasley, Martell L. "Absenteeism and truancy: Risk, protection, and best practice implications for school social workers." Children & Schools 26, no. 2 (2004): 117-128.

Western, Bruce, and Becky Pettit. "Black-White wage inequality, employment rates, and incarceration 1." American Journal of Sociology 111, no. 2 (2005): 553-578.

Appendix A - Variables Selected for Decision Tree Analysis

Socioeconomic Factors	
id	Internal database ID
byear	birth year
ftype311	family type
gender	gender
race	race

School Records	
absen312	Percent absent spring cohort
scsrc311	CTBS reading comp
scsmp311	CTBS math

Psychological Well-being	
pcpsd31f	Which children are sad a lot
pcpwr31f	Which children worry a lot
pcpaf31f	which children are afraid a lot
pcpcr31f	Which children cry a lot
pfmax311	Parents rate child anxious symptoms
pfmdp311	Parents rate child depressive symptoms
pccax311	mean child report anxious symptoms intervention
pccdp311	Mean child report depressive symptoms intervention

Social Adaptation Status	
Peer Assessments	
scpag311	Mean % peer nomination aggression score
scpli311	Mean % peer nomination likability
scpsb311	Mean % peer nomination shy
scpsp311	Social preference score % mean nominations for likability vs do not like
Attention/Concentration Problems, Hyperactivity, & Impulsivity	
sctop311	Teacher child overall school progress
sctob311	Teach child overall classroom behavior
sctcp311	Mean teach rated concentration problems
scthy311	Mean teacher rated hyperactivity
sctim311	Mean teach rated impulsivity
sfmcp311	Parents rate child hyperactivity family
sfmim311	Parents rate child impulsivity family
scted311	Mean teacher rated educational performance
Conduct Problems/Oppositional Defiant Behavior	
sctod311	Mean teacher rated oppositional - defiant behavior
sctcd311	Mean teacher rated conduct disorder problems
scttp311	Total problem score
sctaa311	Mean teacher rated authority acceptance
sctdr311	Teach reported disciplinary removal
sfmod311	Parents mean rating of child oppositional defiant
sfmcd311	Parents mean rating of child conduct problems
Social Contact/Shy Behavior/Likability/Rejection	
sfmli311	Parents rate of child likeability
sfmsb311	Parents rate child shy behavior family
sctsb311	Mean teach rated shy behavior
sctli311	Mean teach rated likability

Appendix B - Technical Explanation of Model Selection

A supervised decision tree algorithm will be used to develop a model from the dataset. After running Chi-Squared Automatic Interaction Detector, Logistic Regression, Classification and Regression (C&RT), Neural Network, and C5.0 algorithms, the C&RT algorithm produced the optimal model with the highest accuracy. All models achieved between 60-70% accuracy; however, all models were overfit to the data by about 15-20%.

The C&RT algorithm, a decision tree algorithm that dichotomizes similar subsets of data based on recursive partitioning, yielded the initial highest accuracy. The standard model, shown in Figure 7, produced 72% accuracy.

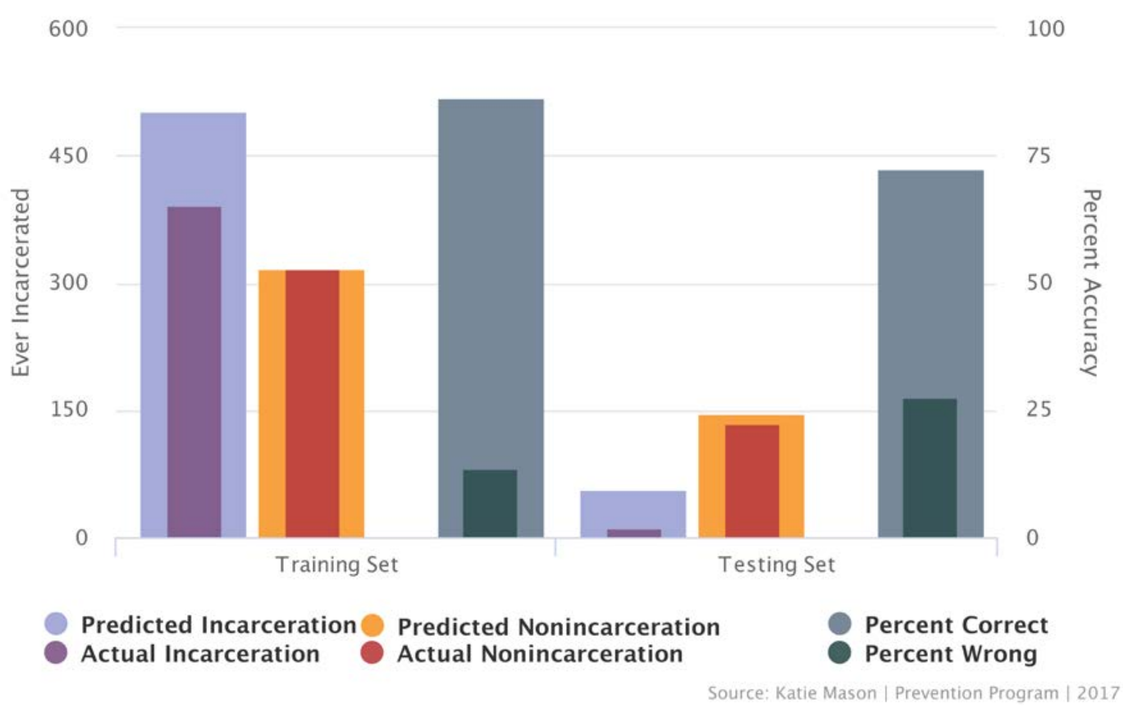


Figure 7. C&RT Accuracy Analysis: Standard Model

This model also provides boosting to improve model performance. The optimal model, shown in Figure 8 was developed with boosting enabled to obtain a more accurate tree model, resulting in 82% accuracy, a 10% increase with only a 3% increase in overfitting.

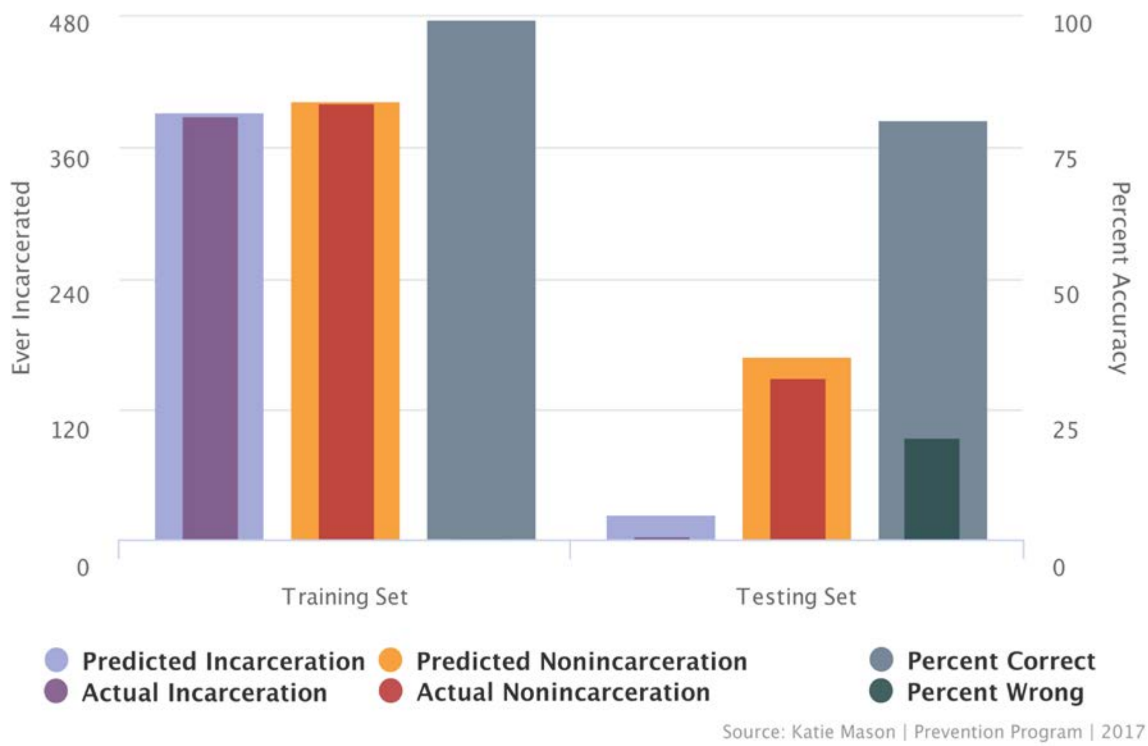


Figure 8. C&RT Accuracy Analysis: Boosted Model

Curriculum Vitae

Kathleen Mason was born in Jacksonville, Florida in December 1982. She is a candidate for Masters of Science in Government Analytics at Johns Hopkins University. She earned a Bachelor of Business Administration with a concentration in Finance from Emory University. After graduating in 2005, Kathleen served as an Ambassadorial Scholar for Rotary International in Dakar, Senegal. Subsequently, she returned to Jacksonville to work in finance as a high net worth securities trader.

More recently, Kathleen transitioned into working in prevention research, most notably with the Johns Hopkins Center for Prevention and Early Intervention on the Prevention Program and the Johns Hopkins Adherence Research Center on the Asthma Basic Care Headstart Program in Baltimore, Maryland. Her research interests include social reform, predictive analytics, and data science.