ARE THOSE TEENAGERS REALLY UP TO NO GOOD? DEVELOPING A PREDICTIVE MODEL OF JUVENILE CRIME

By

Heather Prince
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Criminology, Law, and Society

Committee:

______________________	Director
David B. Wilson	
______________________	
______________________	
______________________	Department Chairperson
______________________	Program Director
______________________	Dean, College of Humanities and Social Sciences

Date: April 25, 2022

Spring Semester 2022
George Mason University
Fairfax, VA

Are Those Teenagers Really Up to No Good? Developing a Predictive Model of Juvenile Crime

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at George Mason University

By

Heather Prince
Master of Science
University of Pennsylvania, 2018
Bachelor of Arts
Albright College, 2017

Director: Cynthia Lum, Professor, Director of Center for Evidence Based Crime Policy, College of Humanities and Social Sciences

Spring Semester 2022
George Mason University
Fairfax, VA

DEDICATION

I dedicate this dissertation to my family and my partner; your support, listening ears, and love have all been invaluable throughout my education and career. Mom and Dad, I can't thank you enough for all the love, encouragement, and support you've given me throughout my life. I couldn't be prouder to be Dr. Heather Prince, and I am so happy I get to share this accomplishment with you. Grandma and Poppops (Rose and Gunther) and Grandma and Poppops (Barbara and Bernie) – thank you for your endless love, support, encouragement, and for believing in me always. Poppops – thank you for building the desk that this dissertation was built on.

Matt, your love and support mean the world to me – I could not be happier to have you by my side and to be sharing life with you. Thank you for keeping me laughing, letting me vent, supporting my stress-snacking, and just being you - the best partner in the world. I love you.

Mr. and Mrs. Deforge, Papa and Tisa – thank you all for your continuous support, encouragement, and love as well. I am so happy to be sharing this time in my life with you all.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the help and support of my fantastic mentors/dissertation committee. Dr. Lum, Dr. Gill, Dr. Wilson, Dr. MacDonald, and a special shout-out to Dr. Ridgeway – you are all the exemplars of what it means to be a scholar and a mentor. I am lucky to have worked with and learned from you all throughout my graduate experience. Dr. Lum, I am thrilled that I will continue to be a part of developing more effective, efficient, and just policing in my career through evidence-based practices and crime policy. You never *really* leave CEBCP. Dr. Gill, thank you for being a fantastic leader and support as we've worked through the true reality of messy data/research. Dr. Wilson, thank you for helping me realize I really do like statistics, and for your priceless machine learning tutorials. Dr. MacDonald, thank you for continuing to mentor and encourage me post-Penn Criminology. Dr. Ridgeway, thank you for your always invaluable R help/sage wisdom/advice – and periodically continuing to help me troubleshoot ~4 years after I've graduated.

Thank you to Dr. Rice, Prof. Abodalo, and Dr. Brown at Albright College – I couldn't have had better mentors; you've all had a hand in putting me on the path that has led me to where I am today. They say hindsight is 20/20, but looking back to the beginning of college, I never would have dreamt that I'd be completing a Ph.D. Thank you for your encouragement, wisdom, and countless hours of wonderful conversations and learning.

To all my friends, thank you for listening to my endless machine learning "thinking out loud", making me laugh every day, being awake at 1:00 a.m. on the group chat while I'm still up working, your encouragement and support, and stress-eating with me. I couldn't be happier to have you all in my life.

I can't have an acknowledgements section without mentioning the pets – Princess Buttercup, thank you for sitting and purring next to/on me while I wrote most of this dissertation. Grendel, thank you for sitting on my lap and keeping me company. Tiger and Proffy, thanks for being with me through everything, and continuing to keep an eye on me.

Finally, thank you to the Fairfax County Police Department for sharing the crime data with me that made this dissertation possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

XGB……………………………………………………………Extreme Gradient Boost
XGBRF…………..………………………………Extreme Gradient Boosted Rando Forest
LASSO……………………..………..Least Absolute Shrinkage and Selection Operator

ABSTRACT

ARE THOSE TEENAGERS REALLY UP TO NO GOOD? DEVELOPING A PREDICTIVE MODEL OF JUVENILE CRIME

Heather Prince, Ph.D.

George Mason University, 2021

Dissertation Chair: Dr. Cynthia Lum

Juveniles in the U.S. (people younger than 18) are believed to be responsible for most crime and often make up a majority of police contacts. However, for most crime that occurs, we do not know who or how old the offender is, and therefore are left to estimate exactly how much crime is actually committed by juveniles, such as self reports or arrests. Police agencies often use crime data to discern "hot spots" of crime, and may assume those crime clusters are primarily committed by juveniles if they occur after or near a school. Yet, such assumptions (and actions based on those assumptions) could be imprecise and inaccurate. Indeed, in the majority of police contacts with young people no crimes or illegal items are discovered. This points to inconsistencies in both knowledge and policy for juvenile offending. To better understand juvenile crime and inform youth prevention policies and practices, this dissertation focuses on creating a predictive model to determine the probability of whether a reported crime was committed by a juvenile. The predictive model usees characteristics of the offense, spatial and temporal elements and other factors to predict whether an offense was committed by a juvenile. Having a more accurate prediction model of juvenile offending could lead to more precisely targeted prevention initiatives, more cost-effective use of police and community resources, and more effective crime control or prevention of juvenile crime. The development and validity of the model are tested using police data from a police department located in a large metropolitan area in the mid-Atlantic United States. The most successful model can predict juvenile status at about 90% accuracy, and findings indicate potential for machine learning to be used for research and understanding unknown patterns of juvenile crime. Practical implications and limitations are also discussed.

CHAPTER 1: INTRODUCTION

Juvenile crime (offenses committed by people under the age of 18) has long been a focus of interest for scholars, policymakers, and law enforcement officials. Significant portions of crime trends in the U.S. can be explained by the proportion of the population belonging to the under 18-year-old age group (Steffensmeier & Harer, 1987; 1999). Given that age strongly correlates with criminal behavior, it is no surprise that young people are believed to be responsible for most crimes and are subject to the majority of police surveillance, stops, or searches (Fratello et al., 2013). Yet, for most of these stops of juveniles no crimes or illegal items are ever discovered (Fratello et al., 2013). This points to a possible disconnect between juvenile crime that actually occurs and police surveillance and enforcement activities toward juveniles.

One reason for this disconnect is that while we infer, given decades of criminological research, that juveniles are responsible for many crimes, we do not actually know this. Most studies of juvenile offending come from self-reports. However, for the daily crimes that are regularly reported to and recorded by the police, there is often no information about the offender, and many crimes go unsolved. This means that the crime data information that police are using to determine crime hot spots to target or to develop youth crime prevention policies, is uninformative when it comes to an offender's age. I call this the "dark figure" of juvenile crime. This dark figure also reflects the biases in how juvenile offending is reported and processed. There are several biases and factors that affect whether an arrest will occur (race, ethnicity, age, gender, socioeconomic circumstances, presence of parents, seriousness of offense, amount of evidence, etc.) that determine whether a juvenile offender is reported to the police, arrested, or officially processed (Liederbach, 2007; Lynch, 2002). Police officers also have discretion when making decisions to make an arrest or not to make an arrest, which leads to biases in arrest data as well (See Liederbach, 2007; Fratello et al., 2013; Bannister, Carter, and Schafer 2001; Giblin, 2002; Withrow and Bolin, 2005).

Current police enforcement efforts on juveniles are “wide net” and assumption-based, and may be inefficient, ineffective, and discriminatory. Juveniles are often suspected of being up to no good, especially when they hang out in unsupervised groups, based on their age, dress, or the locations in which they hang out (Lynch, 2002). Young people are categorized by adults and by police as being either “in place” or “out of place” depending on how they fit into the societal norms of such a space; age, dress, and perceived social background can make youth subject to police surveillance or questioning (Kennelly, 2011). Wearing certain clothing items, for example, a “hoodie” with the hood pulled up, concealing the face, can cause youth to be perceived as being “suspicious” and therefore subject to increased surveillance, but wearing these articles of clothing is simply part of being a youth (Kennelly, 2011).

Such assumptions and lack of information may lead police to target young people in un-strategic ways, based on their age, dress, and location (Black, 1980; Flacks, 2017; Fratello et al., 2013; Lynch, 2002). Such strategies can also be at high risk for being discriminatory based on age and race (see Liederbach, 2007). Additionally, young people do not yet have control of private spaces or own property and are therefore the main users of public spaces, making them more susceptible to police surveillance and discipline (Baumgartner, 1988; Black, 1980; Flacks, 2017; Haller, 1976; Werthman and Piliavin, 1967). Young people under the age of 18 also lack the full advantages of citizenship and adulthood and are more vulnerable during police encounters, feeling less entitled to complain about or challenge police treatment, and feeling as though they are not free to leave when they are legally entitled to (Flacks, 2017; Fratello et al., 2013).

Studies indicate that juveniles indeed account for a significant portion of police contacts with the public (Geistman and Smith, 2007; Walker and Katz, 2008). In one study of youth in New York, Fratello et al. (2013) finds evidence of youth being frequently stopped by police; 44 percent of youth surveyed reported having been stopped nine times or more, 71 percent reported being frisked and 64 percent reported having their clothing or bags searched, while 45 and 46 percent reported experiencing threats or use of force during the stops, respectively. Others have also found that young people are more likely to be subject to use of force by police during such stops (Engel, 2000; Paoline and Terrill, 2005). Fratello et al. also find that despite the heavy surveillance of youth and persistent stop and frisks, 85 percent of youth responded that when stopped, the officers never uncovered any illegal items or activity, and that only 29 percent of respondents were ever given a reason for being stopped. Consequentially, these frequent stops can erode relationships between youth and the police (Brunson and Miller, 2006;

Carr et al., 2007; Friedman, Lurigio, Greenleaf, & Albertson, 2004; Hurst, 2007; Hurst and Frank, 2000; Leiber, Nalla and Farnworth, 1998; Sharp and Atherton, 2007; Taylor et al., 2001; Whitehead and Lab, 1999). This evidence makes it clear that police (at least in New York City where Fratello et al.'s study was conducted) have been using "wide-net" strategies based on assumptions about youth that are made because of a lack of precision in their knowledge about juvenile crime. However, if police knew the "who, where, and when" of juvenile crime, they could theoretically employ more targeted enforcement strategies. Research illustrates that proactive, focused policing strategies have been effective in reducing other types of offending, including gang violence, and crime that occurs in small areas known as "hot spots" (See generally; Braga et al., 2001; Braga et al., 2014; Kochel et al., 2015; Lawton et al., 2005; Mohler et al., 2015; Ratcliffe et al., 2011).

Instead of using "wide-net" strategies that may be inefficient and ineffective, more accurate measures are needed as to where, when, how, and why juvenile crime occurs, as well as how these patterns differ from adults and by different juvenile age groups. This would allow for better identification of places or times that need more effective and precise juvenile crime prevention strategies as opposed to currently used strategies. This dissertation will pilot a method of learning more about the "dark figure" of juvenile crime by working backward, starting with a crime that has been reported and using the elements of that offense that can inform the probability of whether the perpetrator was a juvenile. Developing this predictive model of juvenile offending that can assist in unlocking the “dark figure” of at least reported juvenile crime and expand criminological theory in this area. Such models can also help police departments to better assess whether a crime was committed by a juvenile, and in doing so, better understand spatial and temporal patterns of juvenile crime. In turn, sharper knowledge about juvenile crime could allow both police agencies and communities to develop more proactive, targeted, and justified prevention practices aimed at not only mitigating juvenile crime but also reducing discriminatory or wide-net practices as discussed above.

CHAPTER 2: LITERATURE REVIEW

Criminology theories have long examined juvenile (people under 18) offending, offering a variety of reasons for why, where, when, and why young people commit crime. These theories—and their associated empirical research—provide important clues in developing a predictive model for whether any given crime is committed by a juvenile. Most criminological theory has focused on dispositional explanations for juvenile offending – the "who" and "why" of juvenile crime. However, place-based and routine-activities theories also contribute to the "where" and "when" of juvenile offending. In totality, these studies provide important clues as to whether any given offense that is reported to the police might be committed by a juvenile.

The "What": The Age-Crime Curve

The age crime curve is one of the most well-established theories or explanations of criminal offending in criminology (Farrington, 1986). This visualization of age and offending patterns illustrates that offending tends to rise and then peak in the teenage years, decreasing substantially as people enter late adolescence and adulthood. One of the most common explanations for this desistance from offending is that as people reach adulthood, they have more stakes in conformity (spouses, jobs, college education, etc.) and cannot offend due to unwillingness and/or lack of time (Farrington, 1986). Wilson and Herrnstein (1985) propose that people's ability to delay gratification and consider future consequences for their actions increases with age, which can also cause a decrease or desistance in offending. Indeed, Petersilia et al. (1978) find that common explanations for youth offending (according to self-reports) include excitement and peer influence, and West and Farrington (1977) find similar results. Greenberg (1979; 1983) also emphasizes the role of social institutions (employment, military enlistment, marriage) providing informal social control and contributing to the desistance from crime at older ages. The age crime curve can illustrate age-offending patterns for most youth, though the select few "life-course persistent" offenders will continue to offend beyond youth (Moffitt, 1993). The focus of this dissertation is on those youth that are in the peak ages of offending, (below age 18) during the teenage years. Further, the age crime curve theory holds when examining juvenile offending or risk-taking behaviors through a neurological perspective as well. There is evidence that juvenile risk-taking behaviors (delinquency/offending included) increase during adolescence due to changes during puberty to the brain's socio-emotional system (Steinberg, 2008). This leads to an increase in reward-seeking, especially in the presence of peers, groups of whom often offer informal social rewards such as popularity or positive attention when someone joins them in a risky behavior. Risk taking (offending) declines on the downward slope of the age crime curve as the brain's cognitive control system changes to allow for increased self-regulation (Steinberg, 2008).

The “Who” and “Why”: Social Bonds, Social Learning, and Developmental Theories

Much of juvenile offending is attributed to juveniles “hanging out” in unsupervised peer groups, and with criminally involved peers. This kind of socialization takes place during early and middle adolescence and early adulthood (10 to 25 years old) when youth are more independently exploring peer groups, relationships, and friendships. Some theories suggest that it is these social bonds that drive youths’ propensity to offend, or not to offend (Hirschi, 1969). Social bonds theory (Hirschi, 1969), suggests that the more attached, committed to, involved in, and invested in (believe in) conventional activities and goals one is, the less likely it is that they will offend. This is in part due to social bonds with others that are also attached, committed to, believe in and are involved in conventional activities. Fear of disappointing or angering friends or family members with bad behavior or offending may deter offending behavior (Burgess and Akers, 1966; Hirschi, 1969; Osgood et al., 1996). Research specifically examining familial bonds finds that adolescents who have close relationships with their parents and whose parents closely monitor their activities are less likely to use drugs (Hoeve et al., 2012; Jackson, 2013; Kelly et al., 2011; Mounts, 2002). Being involved in conventional activities also offers social rewards, such as increased social status and recognition from peers. Youth who are strongly connected to their schools or communities, for example, are less likely to engage in drug use (Bryant and Zimmerman, 2002; Dufur et al., 2013; Wray-Lake et al., 2012).

These informal social reward systems also operate within deviant peer groups, where social recognition and popularity can motivate offending (Burgess and Akers, 1996; Hirschi, 1969; Osgood et al., 1996). For example, research shows that those with weakened social bonds are more likely to consume alcohol (Brenner et al., 2011; Ragan et al., 2014; Reyes et al., 2012), and are more likely to engage in use of marijuana or cocaine (Akers and Gang, 1999; Fagan et al., 2013; Hemovich, Lac and Crano, 2011; Perra et al., 2012). Further, it is established that most juvenile offending occurs in groups, usually of two or three persons, who are loosely associated with one another (Reiss and Farrington, 1991). When youth interact and bond with peer groups, whether conventional or deviant, they learn the corresponding behaviors from those groups, and carry out those behaviors when they are interacting with their groups.

Peer groups have a large influence on youth behavior as they begin to mature and separate from their parents and can teach and influence conformity and delinquency. Differential association and social learning theory suggest that being socially involved with deviant peers affords youth the opportunity to learn how to commit crimes and deviant acts, through association with those deviant peers (Akers, 2001; Sutherland, 1947). These theories posit that criminal behavior is learned through interactions with others in intimate personal groups and relationships, the same way that other behaviors are learned (Akers, 2001; Sutherland, 1947). This learning process may include learning techniques necessary to carry out criminal behavior, as well as motives, rationalizations, and justifications for it (Akers, 2001; Sutherland, 1947). Sutherland (1947) specifically posits that when violating the law produces more favorable outcomes than unfavorable ones, then one will choose to offend rather than conform. This is a main component of social learning and differential association theories; criminal behavior is learned and reinforced when the positive consequences of the offending behavior are more powerful than the positive consequences of conforming behavior (Akers, 2001; Burgess and Akers, 1966; Sutherland, 1947). When deviant behavior is rewarded with positive outcomes, the likelihood that one will engage in deviance increases (Akers, 1973; 2001; Burgess and Akers, 1966; Sutherland, 1947). Research illustrates that social learning is significantly related to adolescent drug use (Durkin, Wolfe, and Clarke, 2005; Ford, 2008; Norman and Ford, 2015; Trucco et al., 2011; Vito and Higgins, 2013). However, most youth do not interact with the same peer groups for their entire lives. As youth mature and experience different stages of the life course, they interact and bond with many different peer groups, some of which may conform and some of which may be deviant. As these peer groups change and evolve, so does frequency and prevalence of delinquency and offending.

While these theories may explain why youth might commit crimes, developmental and life-course theories provide greater insight into the link between age and elements of crime (who). For example, we know from the age-crime curve (Farrington, 1986; Moffitt 1993; Robins, 1978) that offending tends to peak in early adulthood, declining quickly afterward. The same is true for group offending; offenses committed by groups of three or more become relatively uncommon after age 20, and those with four or more persons become infrequent much earlier, around age 17 (Lantz and Ruback, 2017; McCord and Conway, 2005; Reiss and Farrington, 1991). This has led to a paradox about those committing crime in youth and adulthood: Robins (1978) found that adult anti-social behavior is closely linked to youth anti-social behavior, but most anti-social youth do not become anti-social adults. Studies find that co-offending decreases with age, further supporting the Robins (1978) paradox (Lantz and Ruback, 2017; Reiss and Farrington, 1991). As offenders become older, the likelihood that they are offending alone increases (Lantz and Ruback, 2017). Reiss and Farrington (1991) find that offenses that were least likely to involve co-offending (violence and fraud) were also most likely to occur at older ages. This may be due to other offenders in the group desisting, leaving only the rare and more serious "life-course persistent" offenders to continue (Carrington, 2002; Conway and McCord, 2002; McCord and Conway, 2005; Moffitt, 1993; Piquero, Farrington, and Blumstein, 2007). Both developmental and life course theories also suggest why crime declines so quickly after adolescence. Offending behaviors tend to decline sharply after adolescence for most youth, as this is a pivotal age where the transition to adulthood begins. Youth may find a partner and get married, start a family, join the military, attend college, or become employed full-time. All of these can take place in early adulthood and are protective factors, barriers, or turning points away from offending behaviors (Laub and Sampson, 1993; Loeber et al., 1998; 1999; Sampson and Laub 1990). In addition to the social events and transitions that can curb youth offending, chemical changes take place in the brain as youth mature that allow for greater self-regulation and ability to delay gratification (Steinberg, 2008). For life-course persistent offenders, offending does not cease, but changes over the life course depending on age and social circumstances (Laub and Sampson, 1993; Loeber et al., 1998; 1999; Moffitt, 1993; Sampson and Laub, 1990; 1992).

Developmental theories also suggest that except for a very small minority of offenders, most juveniles are committing minor offenses. Moffitt argues that for the more common "adolescent-limited" offenders, offending behavior tends to be less severe, less consistent, and dependent on situational benefits. An adolescent limited offender might engage in anti-social behavior in an unsupervised peer group where it would benefit their social status (Hoeben and Weerman, 2014; Moffitt, 1993; Osgood et al., 1996; Rebellon et al., 2019), but may conform to social norms and rules at home or sports practice. This also explains the large drop off in offending when adolescence ends, and adulthood begins. A married or working person trying to develop a family and stay employed would not benefit from offending; offending would cause them harm in their relationships, careers, social status, and finances.

Taken together, these dispositional theories can inform prediction models about whether a reported crime is committed by a juvenile. For example, the likelihood of minor graffiti or property damage being committed by a juvenile is greater than for an adult. Major crimes, such as a burglary-homicide, would be very unlikely to have been committed by a juvenile, given all that we know from juvenile crime theory and patterns of juvenile crime. Juvenile crimes, given the above, tend to be minor crimes of opportunity, often committed in groups, and not very sophisticated. For example, a group of teens drawing vulgar language graffiti on a local playground when they hang out there at night, or defacing a building or sign. These incidents happen when the group of minors is unsupervised, hanging out together, or not doing a structured activity such as playing a sport.

More "Why": Situational and Opportunity Theories

Social control, learning, and developmental theories tend to focus on a juvenile's disposition rather than other circumstances or situations that might contribute to crime offending. In his critique of dispositional theories, Clarke (1980) argued that crime results not simply from an individual's disposition but also because situations and opportunities present themselves for offending to occur. Clarke's (1980; 1983; 1995) opportunity theory (and subsequent situational crime prevention approaches) does not aim to explain crime through specific root causes, but instead focuses on contextual elements of the crime that provide opportunities for offending. Situational and opportunity theories of crime explain that would-be offenders need a specific situation or opportunity structure (see Figure 1, Clarke 1995, p. 103) to commit a crime. These opportunity structures are influenced by both routine activities (see below) and the physical environment, and are characteristics of victims and targets that might attract motivated, would-be offenders. According to Clarke, preventing crime has less to do with fixing dispositions and more to do with addressing opportunities for crimes. For example, situational crime prevention measures might include locking doors, increasing security or surveillance, or redirecting offenders away from targets.

For youth especially, offending may be the result of taking advantage of opportunities, which may also be encouraged by others in unsupervised peer groups and the lack of bonds, as discussed above. These opportunities can arise anytime youth are “hanging out” unsupervised; if everyone in the group is buying spray paint to graffiti the local playground, one may look “uncool” if they decide not to join in. If everyone in the group has swiped a candy bar from the corner store, and one’s crush is in the group waiting for them to do it too, there is peer pressure to join the action, as well as romantic attention to be gained (Osgood et al., 1996; Rebellon et al., 2019). Generally, when the perceived risk of apprehension is greater than the perceived reward of offending, the opportunity to offend, if present, becomes much less desirable (Nagin, 2013). The same holds true for youth – most youth do not offend alone, as there is no reward to be gained from spray painting a swing set by oneself. The rewards are social and peer-based – offending is exciting, fun, and provides social status rewards if peers are around to see it happen, to encourage it, and to join in on the opportunity (Akers, 2001; Osgood et al., 1996; Sutherland, 1947). Opportunity theories can also inform whether a crime is committed by a juvenile, especially when connected to when and where those crimes occur. Many juvenile crimes are crimes of opportunity; they occur when juveniles are unsupervised, and often in groups because there is opportunity for social reward. If there is no one at the local playground, and a group of bored teens is hanging out there, one or more may decide that some spray paint might make for a fun evening; with no one to catch them in the act and the social rewards of acting out in front of peers at stake, the opportunity to offend presents itself. Such theories of the “when” and “where” are next explored.

The “When” (and also “Why”): Routine Activities Theory

Routine activities theory (especially when connected to theories of opportunity) is also important in predicting whether crimes are committed by young people and how juvenile crime patterns emerge. In particular, Clarke (1995) suggests routine activities and lifestyles influence the crime opportunity structures that people encounter. Cohen and Felson (1979) suggest that crime occurs when the convergence of three elements—which can be influenced by macro social forces—occurs: a motivated offender, a suitable target, and the absence of capable guardians. All three of these elements may converge when youth are together in unsupervised peer groups, explaining why crime and delinquency so often occur when juveniles are unsupervised and partaking in unstructured activity together (Brantingham & Brantingham, 1993; Hoeben and Weerman, 2014; Osgood et al., 1996). Osgood et al. (1996), for example, apply routine activities theory to juveniles at the individual level, and replace Cohen and Felson’s (1979) “motivated offender” concept with an assumption that the motivation lies in the deviant act itself; the easier the act and the greater the tangible and symbolic rewards, the greater the motivation to commit deviance. Deviance is more likely to occur when unsupervised groups are just “hanging out” rather than when they are doing a structured activity, such as going on a date or playing a sport (Hoeben and Weerman, 2014; Osgood et al., 1996). For example, research finds that adolescent substance use takes place during unstructured activities, when peers are present, and when authority figures are absent (de Jong, Bernasco, and Lammers, 2019).

In contrast to their time spent unsupervised, juveniles often have very structured routine activities; they are required to be in certain places at certain times almost every day (for example, like school during the day, and home during the night). This means that their offending might be more restricted to certain times and places than adults. In addition, juvenile routes that are traveled to and from these routine activities also tend to stay the same day to day (Brantingham and Brantingham, 1993). Juvenile offending may take place along these routes to and from routine activities; for example, property crimes tend to concentrate near the offenders' places of routine activity and along their normal routes to and from those places (Brantingham and Brantingham, 1993; Rengert and Wasilchick, 1985). The Office of Juvenile Justice and Delinquency Prevention (OJJDP) (2014) report on juvenile victims and offenders finds that violent crimes committed by juveniles peak between 3 and 4 p.m., coinciding with the end of the school day. On non-school days, this number peaks between 6 p.m. and midnight (Sickmund and Puzzanchera, 2014). Similarly, juvenile crime may be more concentrated at particularly places than adult crime. Weisburd et al. (2009), posit that juvenile crime is especially concentrated in certain places that juveniles like to hang out absence of structured supervision, such as school, malls, movie theatres, and routes to and from these locations may be hot spots of juvenile crime (Brantingham and Brantingham, 1993; Osgood et al., 1996; Roncek and Lobosco, 1983; Weisburd et al., 2009; Wilcox 1973).

Routine activities theory informs patterns of youth offending because it focuses on activities and routines of juveniles as a whole, rather than on an individual basis, giving insight to the bigger picture of juvenile crime patterns (more below). This theory suggests that juveniles may offend, or may be presented with good opportunities to offend, because of the patterns and structure of their routine activities, rather than their individual or peer-group decisions. This theory informs the predictive model of this dissertation, in that it suggests that crimes may be more likely committed by juveniles when they occur at certain times and places in which juveniles are most likely engaged in unstructured activities.

The "Where" (and "Why"): Place-Based Theory

Brantingham and Brantingham (1995) and Sherman et al. (1989) linked routine activities more closely to place-based theories of crime, which in turn also can inform a predictive model on whether a reported crime was committed by a juvenile. Place based theories of crime take elements from several theories, including opportunity theory, routine activities, situational theory and even social disorganization theories[1] to explain crime patterns. Place based theories focus on why crime occurs in specific places rather than why certain people commit crimes (Eck and Weisburd, 1995; Weisburd, 2015). Evidence shows that not only is crime concentrated in small places, but those concentrations remain stable across lengthy time spans (Brantingham and Brantingham, 1993; Weisburd et al., 2004; Weisburd and Green, 1995; Weisburd and Mazorelle, 2000).

[1] Early on, Shaw and McKay's (1942) set the stage for place-based theories in their exploration of social disorganization by illustrating that crime is not evenly dispersed in space or time. Social disorganization theory focuses on socio-economic aspects of crime patterns and shifted the focus of criminologists from the individually based "why" of crime, to the location-based spatial and temporal understanding of crime. In their analysis of juvenile arrests as a proxy measure of juvenile crime, Shaw and McKay discovered that youth arrests tended to occur in areas where there was rapid social turnover (may families moving in and out) as well as high poverty rates. Later, several criminologists focused on opportunity, routine activities, and place-based explanations for crime criticized social disorganization theory, suggesting that it was too generalized in its assessment about where crime was actually occurring and that more precise understanding of where crime occurred (Clarke, 1980; Weisburd et al., 1992). Later, Sampson and colleagues would develop this area into theories of collective efficacy (Sampson et al., 1999).

Brantingham and Brantingham (1993) illustrate that crime concentrates near offenders' central activity nodes, such as home, school or work, or leisure places, and along offenders' usual routes to and from these places. Patterns of crime emerge that closely mirror offenders' routine activity patterns in terms of spatial and temporal location (Brantingham and Brantingham, 1993). These concentrations are not only stable across time, but across different locations (Weisburd, 2015). This theory posits that since crime is concentrated, police and crime prevention resources should be concentrated as well.

Place based theory can apply to explaining juvenile offending, as juveniles often gather in specific places due to their routine activities (Brantingham and Brantingham, 1993; Weisburd et al., 2009). These "hot spots" for youth crime might be malls, movie theatres, or other public retail or recreation spaces where youth commonly "hang out" without adult supervision (Brantingham and Brantingham, 1993; Weisburd et al., 2009). It follows from opportunity theory that youth may have more opportunities to offend in these kinds of places and situations. They are with a group of peers, without adult supervision, in busy places where guardians (i.e., security guards, staff, etc.) may not be able to pay attention to everything, which creates opportunities to offend in these spaces and times that may not exist elsewhere (Brantingham and Brantingham, 1993; Clarke, 1995; Cohen and Felson, 1979). Weisburd et al. (2009) argue that juveniles have limited free-activity space and are often required to be in certain places at certain times, such as school or extracurricular activities, and therefore juvenile crimes are probably also occurring within that limited free-activity space. Schools are focal to a juvenile's daily routine, and it was previously thought that the presence of a school in an area increases the probability of violent crimes (Roman, 2002; 2005). However, more recent studies find that there is no evidence that schools alone increase crime (MacDonald, Nicosia, and Ukert, 2018). In specific situations, the closing of socioeconomically disadvantaged schools can cause crime to decrease in certain places (Steinberg, Ukert and MacDonald, 2019). Overall, the opening of a public or charter school appears to have crime reduction effects in the immediate area (MacDonald et al., 2018). The location of a school may have more to do with criminogenic effects than the school itself, as some schools tend to be located in central areas near shopping centers, malls, and large housing developments. The concentration of students at schools and en-route to and from schools serve as focal points of crime because they concentrate both potential offenders and victims within the same space, creating opportunities to offend (Brantingham and Brantingham, 1995; 1993).

Juveniles are often together unsupervised at certain times, and at specific places as well, leading to the formation of spatial patterns of juvenile crime. In terms of these spatial patterns, facilities such as malls and movie theatres that draw youth together from different communities are likely to be crime generators (Bichler, Malm, and Enriquez, 2014; Weisburd et al., 2009). A small number of "magnetic" locations that are popular and widely appealing enable the concentration and interaction of youth that would otherwise not interact with one another (Bichler et al., 2014). Routine exposure to these types of locations, where youth often lack adult supervision, can create situations that cause crime and delinquency (Anderson and Hughes, 2009; Felson, 2006; Felson and Gottfredson, 1984; Osgood and Anderson, 2004). However, these kinds of locations are often routinely patrolled by police or security guards. The intersection of the routines of juveniles and of police/security creates opportunity for crime, and opportunity to be caught. Logistically, juvenile arrests cluster in the areas where these routine activities overlap, creating hot spots of juvenile crime. This leaves gaps in crime data; what is happening in the areas where the police or security do not routinely patrol; where there are no "capable guardians" present? Logically, what follows place-based, routine activities, and opportunity theories is a spatial and temporal pattern of crime, which coincides with times of the day that youth are not in school or participating in extracurricular activities, as discussed above. However, many studies come to these conclusions using juvenile arrest data as a proxy for spatial and temporal patterns of juvenile crime. Given that arrest data only represents a very small proportion of reported crimes, and given the many biases in arrest data of youth, this may not be the best way to understand juvenile crime concentrations. More accurate predictions of whether crimes were committed by juveniles (whether or not an arrest was made) could facilitate more accurate assessment of juvenile crime concentrations and better inform prevention policies towards those concentrations.

The Who, When, Where, and Why: Integrated Theories

Efforts have been made in the criminology literature to integrate both situational and dispositional theories to explain offending behaviors. Some studies combine routine activities and social disorganization theory, arguing that levels of both disorganization and opportunity vary within and between neighborhoods, at the street segment level (Smith et al., 2000; Rice and Smith, 2002; Weisburd et al., 2012). Places where youth and young people gather or places where opportunities are more likely to arise, are often conducive to higher crime rates. For example, the presence of a high school, the number of bars or taverns/lounges, or presence of grocery stores and gas stations on blocks are positively associated with increases in crimes (Bernasco and Block, 2011; Roncek, 2000; Roncek and Maier, 1991). Agnew (2003) presents an integrated theory of strain, control, and social learning designed to explain the peak of offending in adolescence (the age-crime curve). Adolescents are of an age where they are given some (but not all) adult responsibilities, yet have high expectations placed upon them, but are still not able to enjoy or access all the privileges that come with full adulthood (Agnew, 2003). Adolescents are given less adult supervision (Felson, 1998; Larson et al., 1996), increased social and academic demands (Eccles et al., 1996; Greenberg, 1977; Newman et al., 2000), they are expected to participate in a larger, peer-oriented world (Brown, 1990), they experience an increase in the desire for adult privileges (Moffitt, 1993; Moffitt and Harrington, 1996), and have an increased disposition and opportunity to cope with strain in a criminal manner (Agnew, 1997; Compas et al., 2001; Helsen et al., 2000). Agnew (2003) argues that these factors combine and interact with one another to explain the peak in offending at adolescent age, and the subsequent desistance as adolescents reach adulthood. Because juvenile crime is such a complex issue with numerous causes and correlates, integrated perspectives are important to consider when using theory to inform research and policy. Integrated theories may help to better identify time, place, and offense interactions.

Predicting Whether a Crime is Committed by a Juvenile

In sum, theories about juvenile crime provide important clues that can be used to better predict whether a crime was likely to have been committed by a juvenile. For example, social bonds and social learning theories suggest that youth who are involved with deviant peers are more likely to engage in offending behaviors, especially in groups. Thus, crime that involves more than one person may be more likely to involve young people. Developmental and life course theories also suggest that juveniles are likely to commit minor offenses that are opportunity driven and unsophisticated, such as smoking marijuana when a friend's parents are not home, or drinking at a house party (where supervision is often lacking).

Routine activities and place-based theories are perhaps even more informative in trying to determine the probability that a crime is committed by a juvenile. These theories indicate that certain places may attract unsupervised youth and their peer groups at certain times. Thus, crimes more likely to be committed by juveniles are those that may be reported outside of school hours at specific places such as malls, movie theatres, parks, and other recreational areas where youth can gather unsupervised, or even where they might be encouraged to gather. Alternatively, for crimes that occur within school hours, events that occur close to or inside of the school may have higher probabilities of being juvenile related. Sometimes, juveniles find themselves unsupervised in and around schools, the focal point of their routine activities, and crime can increase during the school day and the times immediately before/after school as well (Roman 2002; 2005). The types of crimes that juveniles are likely to commit also make sense in these kinds of places; vandalism, loitering, using drugs, or minor thefts could likely occur at times where juveniles occupy these spaces (Liederbach, 2007; Lynch 2002). These kinds of places are also examples of places that are near juvenile's homes, that they go to frequently, that they know well, and that they are comfortable in, making them more likely to also be places where youth engage in offending behaviors (Brantingham and Brantingham, 1993; Rengert and Wasilchick, 1985).

As discussed in Chapter 1, policies and practices to address juvenile offending, especially those practiced by the police, are not often entirely informed by the above-mentioned criminological theories and empirics. This is not to devalue the experience and knowledge that police build in the field. Rather, police and communities both make assumptions about both juveniles and crime that may not be supported by what is known from science. In turn, this may impact the efficiency, effectiveness, and costs of prevention and enforcement activities (including the cost of reduced police legitimacy, as Fratello et al. 2013 argues). For example, police and the public use perceptions drawn from their experiences or from their knowledge of crime data as a rough measure of juvenile crime. However, this may be problematic because as already mentioned, we do not have a good sense of who is committing crime in official crime data. Juvenile arrest data may also not be the most accurate measure of the patterns of juvenile crime since arrests represent only a very small subset of juvenile crime, and may reflect biases as to whom, what, and where police focus and the high levels of discretion police officers have with arresting juveniles. All of these measurement problems contribute to the "dark figure" of juvenile crime, where it is unknown if, when, and where youth are offending, because those offenses are not discovered and/or not recorded.

Without an accurate understanding of where, when, and if juveniles are offending, developing effective and efficient juvenile crime policy is challenging, if not impossible. Thus, for this dissertation, I use knowledge from these theories of juvenile crime to develop a prediction model of whether a recorded crime was committed by a juvenile. I will be validating the prediction model by testing it against known juvenile offenses (arrests), to determine how well the factors included in the theories and the model predict the juvenile involvement in those offenses. This is a novel method, and one that could help law enforcement practitioners and policymakers develop insight into where and when youth crime is occurring that may have been previously unknown, allowing for implementation of more efficient and effective youth crime enforcement and intervention strategies. This dissertation will determine if building such a model and moving it toward implementation is feasible.

CHAPTER 3: RESEARCH DESIGN

This study will use decision-tree and regression-based methods to attempt to predict a binary outcome: whether a crime that is reported to the police is likely to have been committed by a juvenile. While there are still biases and problems with what types of crimes get reported to the police, this dissertation focuses specifically on better prediction for the reports that we do have using theoretically informed prediction variables.

The data being used for this study come from a police department in a large metropolitan area in the Southeastern United States. Fairfax County, Virginia has a population of 1.14 million people, with a racial makeup of 50% White, 10.6% Black, 20.1% Asian, and 16.5% Latino (U.S. Census Bureau, 2019). Persons under the age of 18 make up 23.3% of the population in this area, and the area has several schools and youth facilities, as well as malls, movie theatres, and shopping centers with youth-targeted activities, such as arcades. The median household income for the area is $124,831, and 70% of the population aged 16 and above reports having a job (U.S. Census Bureau, 2019). The violent crime rate in Fairfax County was 85.90 (per 100,000 people) in 2019, the most recent data available through the Uniform Crime Report (US DOJ FBI, 2019). The property crime rate was 1192.01 (per 100,000 people) in 2019 (U.S. DOJ FBI, 2019). For context, Fairfax County has much lower crime rates than counties of similar size and population. For example, the violent crime rate in Orange County, FL in 2019 was 561.2 (per 100,000), and the property crime rate was 2963.2 (per 100,000) (U.S. DOJ FBI, 2019). The Virginia state violent crime rate was 208.0 in 2019, and the property crime rate was 1642.7 (per 100,000) in 2019 (U.S. DOJ FBI, 2019). The police data for this dissertation are from the years 2018 and 2019, and contain 36,453 and 35,179 arrest, respectively, for a total of 71,632 arrests taking place in the jurisdiction over a two-year timespan. The data used here are composed of arrests, because in order to develop and validate a predictive model, a known age, or at least known juvenile status, is needed. Other variables in the dataset include arrest dates and times, arrest locations, crime types, arrestee age, arrestee gender, arrestee race, felony/misdemeanor classification, violent/non-violent classification, and Fairfax County residency status. Juveniles (individuals under the age of 18) account for 8,893 or 12.4% of these arrests. Because of problems resulting from the COVID-19 pandemic, 2020 data has been omitted from this project.

To create this prediction model of juvenile crime using this arrest data, indicator variables will be identified including time of day, day of week, month, location of incident (proximity to schools, malls, movie theatres), incident classification (felony/misdemeanor) and type of incident (violent crime, property crime, etc.). These indicator variables are drawn directly from theories of juvenile crime and supporting empirical evidence outlined in Chapter 2 and expanded upon below. Then, data will be split into two portions, 20% in one and 80% in the other, using random allocation in R (R Core Team, 2020) for model building and model testing. This split occurs because the model building data (the 20% set) will be used to build and edit the model; it will be "seen" by the predictive model multiple times while coding and programming. The model testing data (the 80% set) will be used to examine the validity of the final model, on a set of data that the model has not yet "seen" to determine if it can accurately predict whether an offense was committed by a juvenile. These datasets will contain only crimes where an offender (and their age) is known. The purpose of this is to ensure that the model can be tested for accuracy before the possibility of moving on to trying to predict the unknown. The theories, and empirical support of those theories above inform the indicator variables that will be used to build this model. Using the theories and empirical evidence as well as real youth crime data allows for the creation and testing of a predictive model where we will know and be able to confirm that the model's predictions are accurate. Two algorithms will be tested and compared to determine the best fit model, including XGboosted random forest, and LASSO (Least Absolute Shrinkage and Selection Operator) logistic regression. The models will be constructed in R (R Core Team, 2020), and k-fold cross validation will be used to determine the best parameters for each. These concepts are explained in detail below.

Indicator Variables

The indicator variables to be used for this study are informed by the theoretical discussion from Chapter 2. Indicator variables therefore will include time of day, day of week, month, location of incident, whether the incident was a felony or misdemeanor, whether the crime was violent, and what type of crime was committed. Time of day and day of week are important to understanding youth crime as youth have strictly set routine activities; they are required to be in certain places at certain times (school, sports, etc.) almost every day. Most of their routine activities include adult supervision and structured activities, and it follows that youth offending will likely occur outside of those structured and supervised activities (Weisburd et al., 2009). For example, juvenile crime may be more likely to occur during the hours of 2:30 to 4 p.m. on school days (Monday to Friday), and from 6 p.m. to midnight on weekends. Month of the year also ties into these relationships, as youth are only in school (in Fairfax County) during September through early June, so the months of June, July, and August may see increases in youth crime as well. The time of day and date variables are included in the dataset, but will be separated out into individual "hour of day," "day of week," and "month of year" variables for this model building and analysis.

Time of day and dates also tie into the location aspects of juvenile offending. Weisburd et al. (2009) find that youth offending is very concentrated at places and across times. Places where youth gather, often unsupervised, during after school hours or on non-school days, are often youth crime "hot spots." These places include malls, movie theatres, shopping centers, and other places of recreation (Weisburd et al., 2009). Though the evidence for schools being crime concentration areas is mixed (Macdonald et al., 2018; Roman 2002; 2005; Gottfredson et al., 2001), Brantingham and Brantingham (1993) find that juvenile crime takes place around central activity nodes, such as schools, malls, arcades, and along the routes to and from these places of activity. MacDonald et al. (2018) suggest that schools themselves may not be crime generators but may tend to be located in areas with other activity centers and provide a concentration of youth, creating opportunity for crime to occur. Incident proximity to schools and proximity to other areas of social leisure (movie theatres, malls, etc.) will be determined using R (R Core Team, 2020) and the Fairfax County Public School district school locations, which are publicly available on the Fairfax County School District website.

Incident proximity to malls, movie theatres, community centers, shopping centers, or schools can give important clues as to who the offender(s) is, and whether they are a juvenile. This information will be included in the data from the provided X and Y coordinates which will be geocoded using R. Proximity to leisure locations, community centers, and schools will be calculated and included in the dataset in binary format. Using the Leaflet package in R (Graul, 2016), the X and Y coordinates of the arrests will be geocoded and plotted as points. The locations of the schools, community centers, and leisure sites will also be geocoded and plotted as points. A half-mile buffer was added around each point for schools, community centers, and leisure activities (movie theatres, community centers, arcades, etc.). This decision was made because in Fairfax County, students are required to walk to school, or a bus stop, up to 1.5 miles (https://www.fcps.edu/resources/safety-and-transportation). Buffers of 1.5 miles to illustrate walking distances were expansive within the urban layout of Fairfax County, and arrests would have likely all been inside all three buffers. This does not allow for differentiation if an arrest occurred closer to a school than a community center or leisure site, for example; it would appear as the arrest is within all three buffers. The half mile buffers allowed this project to consider some walking distance, as offenses can occur when juveniles are unsupervised on their way to or from these places, but also allows for differentiation between locations where arrests occur. If an arrest takes place inside the buffer for a school, community center, leisure site, or any combination of the three, the corresponding variable in the dataset (*nearSchool, nearComm, nearLeisure*) is coded as 1, while points outside of the buffers are coded as 0. As the incidents recorded get closer to these kinds of locations, the probability that a juvenile is responsible for the offense increases, especially when these crimes occur outside of school hours or in groups.

Crime type and seriousness can also be indicative of an offender's age. Variables that measure crime type/seriousness include whether a crime is a felony, misdemeanor, violent, or non-violent. Additionally, type of crime (i.e., theft, shoplifting, vandalism, assault, etc.) will also be included, as youth tend to be responsible for property crimes and low-level offenses such as loitering, nuisance crimes, drug use and possession, vandalism, and small theft (Liederbach, 2007; Lynch 2002). These kinds of crimes take place when youth "hang out" in unsupervised groups, during times of day or of the year when they are not in school, and at locations where they are not supervised by parents, teachers, or coaches (Bichler et al., 2014; Osgood et al., 1996; Weisburd et al., 2009). Most of these crimes are nonviolent misdemeanors, and generally encompass what most might think of as "teenage mischief," and type of crime, violence, and felony/misdemeanor classification can be important clues as to the age of an offender. Incident type, violence, and felony/misdemeanor classifications will be included in the data, as non-violent misdemeanor crimes of low seriousness are likely to have been committed by juveniles. Interactions between these indicator variables are also important to consider, especially for places and times. Juvenile routine activities are very structured, so they generally are only in unsupervised groups at certain places during certain times, giving a certain opportunity window for offending to occur. For example, if a shoplifting offense occurs at a shopping mall at 11:00 a.m. on a Tuesday in October, the likelihood that a juvenile or juveniles committed that crime is reasonably low, as juveniles are in school, or most should be, at that time. In the data, this would look like: *hour11* = 1, *dayTue* = 1, *monthOct* = 1, *nearLeisure* = 1. Conversely, an incident that occurs at an arcade on a weekend evening in the summertime has a reasonably high likelihood that juveniles were involved. The variables in the data that are being used for the machine learning algorithms must all be in numeric format, and this is elaborated on below. In table 1 below, the variable formats are further illustrated for clarity.

Table 1: Indicator Variables and Corresponding Binary Variables

Variable	Binary Variable Description
Time of Day	*hour1* = 1/0, *hour2* = 1/0, *hour3* = 1/0…through hour 24 (police data uses 24H military time; midnight (0) is the reference category)
Day of Week	*dayMon* = 1/0, *dayTue* = 1/0, *dayWed* = 1/0…through Saturday (Sunday is the reference category)
Month	*mFeb* = 1/0, *mMar* = 1/0, *mApr* = 1/0…through December. (January is the reference category)
Location of Crime	*nearSchool* = 1/0, *nearLeisure* = 1/0, *nearComm* = 1/0 (based on arrests falling within ½ mi. buffer zones)
Felony/Misdemeanor	*FM_dummy* = 1/0, where 0 = misdemeanor and 1 = felony
Violent/Non-Violent	*Vio_noVio* = 1/0, where 0 = non-violent and 1 = violent
Type of Crime	*Typecrime* = 1/0 (i.e., *theft* = 1/0, *drugs* = 1/0, *vehicle_theft* = 1/0) for each crime type in the data

Decision Tree Modeling

Decision tree modeling is an effective way to make predictions, and it allows for more complex variable relationships than linear regression does. In Figure 1 below, an example of a simple, small decision tree is illustrated. Each leaf node depends upon the results of the previous node, and the final prediction that the model calculates is an interaction between each of those variables. In the example below, the prediction of "likely juvenile perpetrator(s)" is the result of an interaction between "incident occurs < 2 miles from school" and "incident occurs between 3-4 p.m."

To prepare the data for XGboosted random forest models, first, the completed dataset will be split into the building (20%) and testing (80%) portions randomly, using random assignment in R (R Core Team, 2020). This technique takes a random sample equal to 20 percent of the whole dataset, and a random sample equal to 80 percent of the dataset (not repeated/overlapping with the 20%), and stores them in two separate datasets: training (20%) and testing (80%). Twenty percent of the crime incident data will be used to build and teach the model, and the other unseen eighty percent will be used to evaluate model accuracy. The data used for building and testing the models will include only those offenses for which an offender and their age is known; this way, the models can be tested and evaluated accurately to ensure that they make predictions correctly. As shown in the example below, each root node in the decision tree represents a single input variable (x) and a split point on the variable. The leaf nodes on the tree contain an output variable (y) that is used to make a prediction.

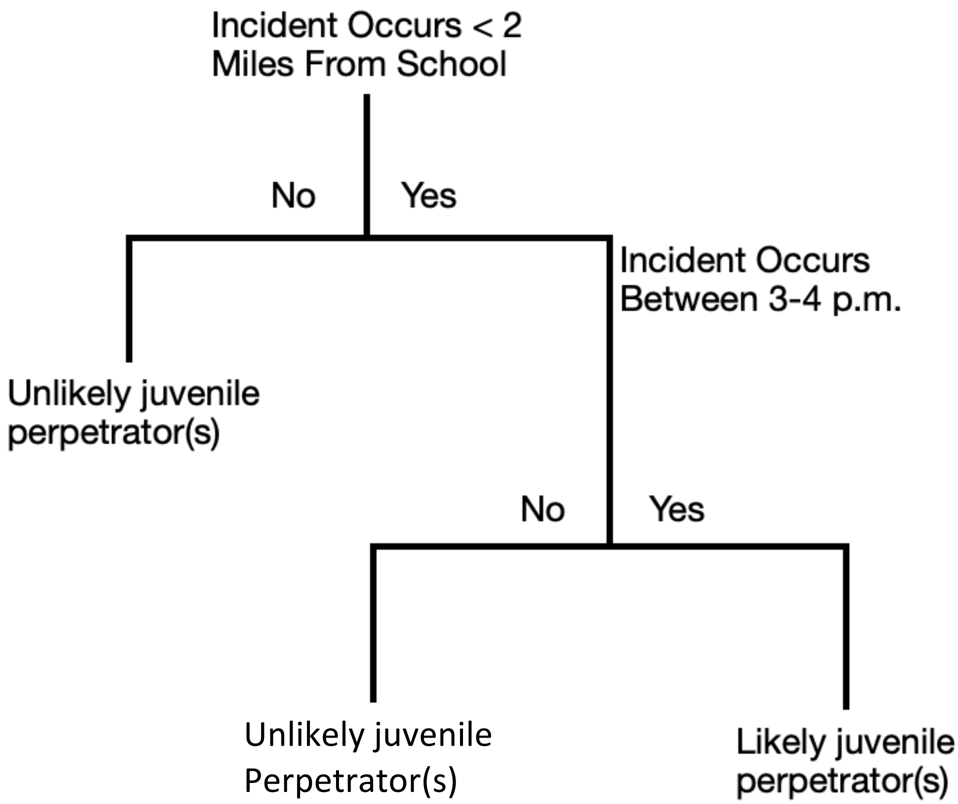

Figure 1: Example of a Decision Tree

In the above example, a simple decision tree is demonstrated. This figure is not based on data nor is it an actual model; it is purely for a simple example illustrative purpose. The decision tree can be stored to a file in R as a set of rules, for example: incident occurs < 2 miles from school: yes; Incident occurs between 3-4 p.m.: yes; therefore: juvenile perpetrator likely. The input variables will be binary numeric variables as numeric values are essential for the XGboost package in R (Chen, He, Benesty, Khotilovich, Tang and Cho, 2015). As illustrated in Table 1 above, time of day will be numeric 24-hour military time, each hour dummy coded with midnight (0), being the reference. Day of week will be dummy coded into individual variables for each day, with Sunday being the reference. Month of year will be dummy coded the same way, with January being the reference. Proximity to schools will be binary with less than 0.5 miles being 1 and more than 0.5 miles being 0, proximity to community centers will follow the same assignment, as well as proximity to leisure sites. Felony/misdemeanor classification will be assigned 1 for misdemeanor and 0 for felony, violent/non-violent crime will be assigned 1 for non-violent and 0 for violent. Finally, crime types (theft, robbery, etc.) will be dummy coded for each type of offense. Table 1.2 below illustrates how race (reference: white), arrest type (reference: summons), sex (reference: male), and county residency status (reference: non-resident) are all also coded as dummy variables. Arrest types included summons and "on view," where summons arrests indicated that a summons was issued and a suspect turned themselves in, and "on view" arrests were arrests made in person on the street, during a traffic stop, etc. Though these are not central to the hypotheses in this dissertation, it is good practice in machine learning to keep all variables and eliminate un-needed ones as model parameters are tuned.

Table 1.2: Demographic/Arrest Specific Variable Coding

Variable	Binary Variable(s) Description
Race	White = excluded as reference Black = 1/0 Hispanic = 1/0 Asian = 1/0 Indian = 1/0
Sex	1 = male 0 = female
Arrest Type	On View = 1/0 Taken into Custody = 1/0 Summons = excluded as reference
County Residency Status	0 = non-resident 1 = resident

All of the indicator variables and their possible split points are evaluated and chosen to select the best-fitting split points possible. To prevent over-fitting of the model, the decision tree is kept as simple as possible, including the lowest number of splits possible. These parameters, among others, are determined with model cross validation, discussed further below, during the "training" stage with the training portion of the data.

Machine Learning and Decision Tree Models

I will be testing and comparing the results from two machine learning algorithms to find the best fit model for the goals of this dissertation, including XGboosted random forest, and LASSO logistic regression. XGboost (Chen and Guestrin, 2016) stands for extreme gradient boosting and uses regularization to avoid over-fitting, giving it better performance (Chen and Guestrin, 2016). Regularization is the process of tuning the level of model complexity such that the model is better at making predictions. For example, predicting whether an animal is a dog or a cat (based on variables like fur length, vocalizations, size, weight, etc.) would not need a very complex model. However, predicting whether it is going to snow tomorrow depends on much more complex variables and interactions, such as pressure systems, storm systems, wind, temperatures, location, etc., and would likely need a very large, complex model to make accurate predictions. Regularization takes place through cross-validation and the tuning of model parameters, detailed below. Random forest models (Ho, 1995) are made up of several (or many) small decision trees that each make a class prediction, and taken all together, those predictions will make up the overall prediction for the full model. However, each individual decision tree must not be too correlated with the other individual trees. The XGB random forest model uses decisions trees that are not independent. The trees are iterative, meaning that as new tress get added to the forest, the model attempts to correct error from previous trees; new trees will try to make better predictions on the incorrect ones from the previous tree. This group of not-so-great-performing small trees, together, create a strong predictive model. (Ho, 1995).

There are several parameters to be tuned once a random forest parameter search has run, including "*nrounds*," "*max depth,*" "*eta*," and "*lambda*." The specifics of the parameters for random forest models are as follows:

- *nrounds* is the number of decision trees in each forest
- *max depth* is the maximum number of decision nodes/branches the decision tree(s) have

- *eta* is the size of the deviation between iterations, or the rate at which the model learns patterns in the data. The default is 0.3, and the entire range is 0-1. Typically, eta lies between 0.1 to 0.3. Gradient Boosting (Chen et al., 2015) involves creating and adding decision trees to a forest sequentially; new trees are created to correct errors in the previous trees. However, due to the nature of random forest models (having several decision trees put together to form a larger, complex model) XGB is prone to over-fitting. *Eta* reins the model in; instead of adding predictions of new trees to the forest with full weight, *eta* is multiplied by the residuals being added, reducing their weight. This reduces model complexity overall, and helps to prevent over-fitting.
- *lambda* is used for regularization – this number tells the model how much to reduce coefficients to avoid over- or under-fitting. The larger lambda is, the more coefficients are shrunk toward zero. If *lambda* is too large, the model will be simple, but likely under-fit. If *lambda* is too small, the model will be more complex, but likely over-fit.

To begin, a parameter search is performed; a range of numbers is set for each parameter and simulations using every combination of the numbers within those ranges are performed on the training dataset. The values of these parameters are data-dependent; the initial parameter search will feature wide ranges of parameters that are narrowed down to more precise values. The result of a parameter search (grid search) will be a list of the optimal parameters for that data, given the ranges that were set to search. If the *eta* or *lambda* values are between the two extremes of the set ranges, those values should work well in the testing model. If the values are at, or close to, one of the extremes in the ranges, it is best to expand the range in that direction and run the search again. For example, if the search range for *eta* was set for 0.3 to 0.8, and the simulation returns an *eta* of 0.7, the search range for *eta* should be expanded in that direction (maybe from 0.3 to 0.95) and run again to see if a more optimal setting for *eta* can be found. Alternatively, if *eta* were to come back at 0.6, this value would likely be a good parameter setting for *eta* in the final model. In an ideal world where money, data storage, and time are infinite, a parameter search including every possible parameter value could be conducted. This also may be possible on an extremely small dataset. However, as datasets grow larger, the amount of time it takes to run a parameter grid search (train the model) grows exponentially. For this reason, search parameter ranges that are relatively large, but not extremely large (ex., a range of 10-15 values), are used to conduct the initial parameter search.

The number of cross-validation iterations also needs to be set; this is how the data will be divided for each set of the parameters set in the ranges. The data is divided into sets based on the number of iterations; during the parameter search, the algorithm will fit models using parameters from the ranges set. Cross validation occurs based on the number of iterations set; for ten iterations (general good practice), the data are divided up into ten sets for every single combination of parameters from the specified ranges, and models are run with one of the ten sets being used as the test set, and the rest as training sets. The model is trained on the training sets and scored on the test set, and this process repeats until each unique group has been used as the test set. The parameters that produce the model with the lowest mean squared error are the ones that will be listed in the result and are the ones that should be set when setting parameters for the final model(s) to be used on the testing data.

LASSO Logistic Regression Modeling

LASSO (least absolute shrinkage and selection operator) regression (Tibshirani, 1996) uses regularization to reduce over fitting and assist in feature selection. Regularization for LASSO models still serves the same purpose; tuning model complexity to create a better fit model. In LASSO logistic regression, the coefficients of the less contributive variables are forced to be exactly zero, and only the most contributive ones are kept in the final model. Unlike random forest, LASSO only has one parameter that needs to be tuned: *lambda*. *Lambda* is used the same way in the LASSO model as it is in random forest; for regularization to avoid over-fitting. This value indicates how much coefficients should be reduced or moved toward zero to optimize the LASSO model. This weeds out variables that do not contribute to the model and improves model running speed, and model performance in making predictions. The simulation run on the training data for LASSO uses cross validation to determine the lowest possible value for *lambda*, *lambda.min*, which should produce the best fitting model for the data. *Lambda* can also be set to *lambda.1se*, which is the largest value of *lambda* where error is still within one standard deviation of the cross validated errors for *lambda.min*. It is generally considered good practice to use *lambda.1se* because using the *lambda.min* value generally produces models that overfit (Krstajic et al., 2014). The general logic of using *lambda.1se* is that it is the simplest model whose accuracy compares to the best model; this acknowledges that regressions are estimated with error, and *lambda.1se* errs on the side of caution; reducing model coefficients, but not by too much.

Putting it All Together

To summarize, machine learning provides a possible avenue to develop a prediction model that could, theoretically, predict whether a crime was committed by a juvenile. XGboosted random forest models can make predictions about non-linear relationships that are complex and involve variable interactions. Random forest modelling uses regularization, the tuning of parameters, to avoid model over-fitting. These parameters include

- *nrounds*, the number of decision trees in each "forest"
- *max depth*, the maximum number of nodes (splits) allowed on a single tree, from the root (start) of the tree to the farthest leaf (last node)
- *eta*, the size of the deviation between iterations, or learning rate of the model. Controls reduction of weight of predictions from trees being added to random forests, to avoid over-fitting.
- *lambda*, the value by which coefficients should be reduced to avoid model over- or under-fitting

The parameters set for each random forest model affect model accuracy and performance, and parameter searches are conducted to ensure that the best possible parameters are selected for use in the models. LASSO logistic regression models can make predictions based on regression modelling techniques and linear relationships, using the tuning (regularization) of *lambda* (the value that coefficients should be reduced) to avoid model over- or under-fitting. Both models can produce numeric probability predictions, that can be converted into binary yes/no predictions.

Though this dissertation is a very early pilot of using machine learning techniques being used to predict the juvenile status of offenders based on characteristics of incidents, using data where ages of arrestees are known (for model validation), a predictive model such as this has the potential to be used in the future by police departments and data analysts. Predictive modelling may allow patterns of juvenile crime to be more accurately captured and can open the door for evidence based juvenile crime policy changes, for example, increased patrol in certain places (e.g., outside the movie theatre) at certain times (e.g., on weekend nights). If the spatial and temporal patterns of juvenile offending are better understood, police can tailor their crime control efforts to the evidence provided by the data, rather than generalizing about juveniles and offending. However, there are several important caveats to be considered when thinking about implementing a machine learning model to predict offender juvenile status in a practitioner setting, discussed in further detail below in the discussion chapter.

CHAPTER 4: IMPLEMENTATION AND RESULTS

General Demographics and Descriptive Statistics

To prepare the data for use in machine learning algorithms, all the variables in the data must be numeric values, as is general standard practice for statistical analyses. In addition, none of the variable values can be null or missing, meaning that all cases with null/missing values needed to be removed. Fortunately, there were only 2648 of these cases in the Fairfax County data. After the elimination of cases with missing values, the total dataset contained 68,984 observations, each representing an arrest made in the county for 2018-2019. Table 1.4 below displays the descriptive statistics for these data.

Table 1.3: Descriptives of Arrests of Juveniles, Fairfax County, 2018-2019

Variable	Categories	Percent of Arrests/Arrestees	Number of Arrests
Sex	Male Female	60.35% 39.65%	5213 3425
Race	White Black Asian Hispanic Indian	65.9% 30.35% 3.67% .0002% .0002%	5693 2622 317 2 2
Residency Status of Fairfax County	Resident Non-Resident	81.88% 18.11%	7073 1565

Day of the Week	Sunday Monday Tuesday Wednesday Thursday Friday Saturday	10.7% 14.4% 15.7% 14.6% 15.9% 16.4% 12.2%	927 1247 1357 1261 1372 1419 1055
Month of Year	January February March April May June July August September October November December	8.62% 7.44% 8.22% 8.49% 8.93% 8.36% 9.09% 6.46% 7.23% 7.54% 10.59% 10.19%	745 643 710 733 771 722 685 558 625 651 915 880
Year	2018 2019	55.1% 44.9%	4760 3878
Misdemeanor/Felony	Misdemeanor Felony	82.46% 17.54%	7123 1515
Violent/Non-Violent Crime	Violent Non-Violent	11.27% 88.72%	974 7664
Most Common Offenses	Theft Drug Offenses Assault Trespassing Destruction of Property/Vandalism Liquor Law Violations Disorderly Conduct Weapons Offenses (possession, weapons in public, etc.) Burglary Motor Vehicle Theft	25.9% 13.5% 8.75% 4.65% 3.02% 2.80% 1.84% 1.11% 0.98% 0.92%	2245 1165 756 402 261 242 159 96 85 80

Of these arrests, 8,638 (12.52%) were arrests of juveniles. Of the 8,638 juvenile arrests, 5,213 (60.35 %) were of males and 3,425 (39.65 %) were of females. In terms of race, 5,693 (65.9%) of the juveniles were identified as white, 2,622 (30.35%) were identified as black, 2 (.0002%) were identified as Hispanic, 2 (.0002%) were identified as Indian, and 317 (3.67%) were identified as Asian. These categories come directly from the Fairfax County police data. Surprisingly, arrests were generally evenly spread throughout the days of the week, with 927 on Sundays, 1,247 on Mondays, 1,357 on Tuesdays, 1,261 on Wednesdays, 1,372 on Thursdays, 1,419 on Fridays, and 1,055 on Saturdays, though the slight uptick in arrests on Fridays is notable.

All 8,638 of the juvenile arrests occurred in at least one of the school, community center, or leisure site buffers. There were 724 overlaps, or arrests that occurred in more than one type of buffer (e.g., when a school and leisure site buffer overlap/intersect). This may be due to several factors, such as county size and development; Fairfax is extremely developed, with schools, community centers, and leisure sites (malls, shopping centers, theatres, arcades, etc.) all relatively close to each other. This ensures that these amenities are conveniently available in a short distance to the many residents that live in the high-density apartment and townhome complexes located throughout the county. Of the juveniles that were arrested, 7073 (81.88%) were identified as residents of Fairfax County, while 1,565 (18.11%) were not residents. This is consistent with general findings in the literature that posit youth often do not travel far to commit crimes, whether that is because of a lack of transportation, or because youth generally feel more comfortable offending in familiar places that they know well (Brantingham and Brantingham, 1993; Rengert and Wasilchick, 1985).

Regarding the seriousness of offenses, the data reflect the general findings that juvenile offending tends to consist of non-violent misdemeanor offenses (Hoeben and Weerman, 2014; Moffitt, 1993; Osgood et al., 1996; Rebellon et al., 2019). The majority of youth arrests were misdemeanors (7,123, or 82.46%) with only 1,515 (17.54%) being felonies. The same pattern followed for violent crimes; only 974 (11.27%) arrests were for violent crimes, while the remaining 7,664 (88.72%) were for non-violent offenses. Unfortunately, due to the nature of the arrest data not including the number of arrestees involved in an incident, it is not possible in this paper to determine proportions for "group" vs. "solo" youth offending. The most common types of offenses that juveniles in Fairfax County were arrested for were theft (2,245) and drug offenses (1,165). This pattern reflects what the literature suggests about the seriousness of most youth offenses: these types of offenses are non-violent and are usually charged as misdemeanors for juveniles. However, felony/misdemeanor classification can also depend on factors such as whether the incident was a first-time offense, or the type and/or quantity of drugs in drug offenses. In the following sections, the ability of these variables to predict whether an offender is a juvenile is explored.

Preparing to Run XGBoosted Random Forest Models

This analysis involves constructing, tuning, and testing the predictive validity of XGboosted random forest models for predicting the age of offenders based on characteristics of an incident. XGboosted random forests are generally a good way to make binary predictions that may involve complex variable relationships and interactions. To begin, the dataset was prepared by converting all the variables to a numeric binary format, as discussed above. Then, the dataset was split into two (20/80) sections using a function in R that allows for random selection of observations without replacement. The “train” dataset is the smaller portion of the data (20%), and the models are evaluated using the larger “test” part of the data (80%).

In the test data (80%), the outcome variable (juvenile Y/N) is set to “null” and the model uses the rest of the variables to make predictions. The outcome variable (juvenile Y/N) is then added back to the dataset after predictions are made, so that the model predictions can be evaluated against the actual observed values, to determine model accuracy. This comparison is known as a “confusion matrix,” which compares the actual values to the predicted ones in a 2x2 table format, discussed further below.

Performing a Parameter Search to Determine Optimal Parameters

The next step in building a random forest model is to set the parameters for the grid search of the parameter space. This grid search performs multiple random forest simulations using all possible combinations of parameters from ranges that are manually set. The optimal parameters for the model from the set ranges are returned as results. For random forest models, the parameters include *nrounds, max depth, eta,* and *lambda*. Finally, the resampling strategy, in this case, cross-validation, is defined. The cross-validation was set for 10 iterations, as that is generally considered standard good practice. This process is important to avoid model overfitting, which can lead to a predictive model that is accurate for the data that was used to build it but fails when presented with new data. In table 2.1, the parameters that were set for the initial grid search, and the optimized parameters that the grid search returned are displayed.

Table 2.1 Parameter Search for Random Forest Model 1

Parameter Name	Parameter Range Searched	Parameter Value Returned
Nrounds	15 to 40	37
Max Depth	2 to 25	22
Eta	.2, .25, .3, .35, .4, .45, .5, .55, .6, .65	0.5
Lambda	1.2, 1.4, 1.6, 1.8, 2, 2.3, 2.6	2.3

For this parameter search, all four of the parameters came back being close to the upper end of the range, indicating that the search ranges needed to be expanded in that direction. This model was not applied to the testing data for this reason; the predictions would not have been very accurate given the untuned parameters. To improve upon the first parameter search, the ranges for the parameter search for model 2 were extended in the appropriate directions after learning where the parameters fell in the first search. Table 2.2 below shows the parameters searched for random forest model 2, and the optimized parameters that were returned from the grid search.

Table 2.2 Parameter Search for Random Forest Model 2

Parameter Name	Parameter Range Searched	Parameter Value Returned
Nrounds	25 to 45	36
Max Depth	15 to 40	28
Eta	.4, .45, .5, .55, .6, .65, .7	0.5
Lambda	2, 2.3, 2.5, 2.7, 3, 3.1, 3.3, 3.5	3.5

For this parameter search, *lambda* came back at the highest end of the searched range. For this model, the optimal number of trees is 36, the maximum depth of trees is 28, the space between iterations, or model learning rate, is 0.5. For *lambda,* the larger the value, the more coefficients are pushed toward zero; however, some coefficients may remain large depending upon their contributions to the model. These parameters were used in an XGB random forest model to make predictions in the test data, and another model was prepared using a different parameter search as well. I performed multiple parameter searches and ran different random forest models due to the exploratory nature of this dissertation; I attempted to improve the tuned parameters for each model, and I also remove unimportant variables to determine if that would improve model performance.

XGB Random Forest Model 2 Performance

Figure 1.1 displays a plot of an importance matrix for model 2. The importance matrix displays variables used in the model and several metrics, but for these models the “gain” metric is most relevant. This metric shows the relative contribution to the model of each variable such that if one variable has a higher value of gain compared to another, that variable is more important for generating a prediction. Gain is the improvement in accuracy brought by a variable to the branches of the decision tree that it is on. In this model, the importance matrix plot in figure 1.1 below illustrates that the five most important variables are; whether it was a felony or misdemeanor, whether it occurred in 2018, whether the arrest was on view, whether the suspect was male or female, and whether the suspect was a resident of Fairfax County. Additionally, whether the incident was a theft, or near a school, community center, or leisure location were in the top ten most important variables.

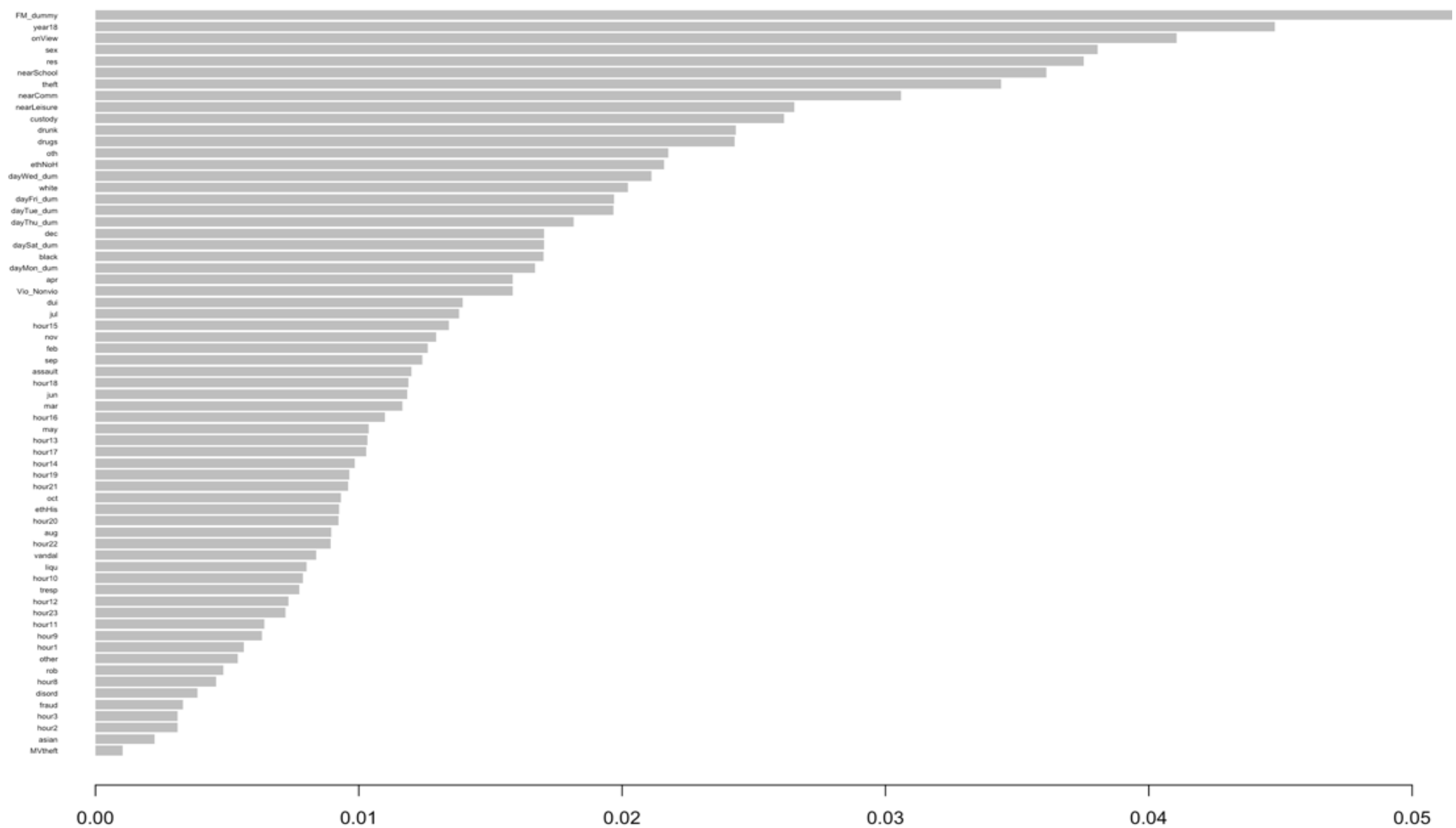

Figure 1.1 Plot of Importance Matrix for XGB Random Forest Model 2

In terms of predictions, the XGB random forest model (XGBRF) produces a column of probabilities in the test (80%) dataset. To obtain binary predictions for each of the observations, a new variable was created where "juvenile" was equal to 1 if the probability was more than 0.5, as is general practice, and equal to 0 if the probability was less than 0.5. The predicted values were then compared to the actual juvenile indicator variable in a confusion matrix to evaluate how well the model was able to predict juvenile offense perpetration. A confusion matrix is a 2x2 table comparing predicted to observed values of a variable, and is used to make predictions for binary classifiers, or predictions that are binary 0/1 (N/Y) predictions. There are several values that are computed from a confusion matrix for a binary classifier, including accuracy, misclassification rate, true positive rate, false positive rate, true negative rate, precision, prevalence, null error rate, Cohen's kappa, and an F-score.

Table 2.3 Confusion Matrix for XGBRF Model 2

	Actual "no"	Actual "yes"
Predicted "no"	46571	4309
Predicted "yes"	1642	2687

The accuracy for XBGRF model 2 returned a value of 0.8922096, or about 89%; indicating that the classifier is correct in its prediction about 89% of the time. In plain English, this means that by using these indicator variables, we can accurately predict whether a crime that resulted in an arrest is committed by a juvenile 89% of the time. To follow, the misclassification or error rate is 0.1077904; the model is wrong about 10.7% of the time. The true positive rate for this model is 0.3840766; this model predicts "yes" about 38.4% of the time that the correct value is actually yes. In other words, this model accurately predicted "yes" on 38.4% of the actual observed "yes" values in the dataset. The false positive rate is 0.0340572; this model predicts "yes" when the correct classification is "no" about 3.4% of the time. The true negative rate for this model is 0.966198, meaning that about 96.6% of the time, this model predicts "no" where the correct value is actually "no." This model accurately predicted "no" on 96.6% of the actual observed "no" values in the dataset. The precision rate for this model was 0.6206976; when this model predicts "yes," it is correct about 62% of the time. Prevalence refers to how often the "yes" condition actually occurs in the sample, and that is 0.1267185, or about 12.7%.

The null error rate identifies how often the model would be incorrect if it always predicted the majority class; in this case, how often this model would be wrong if it always predicted "no." The null error rate for this model was 0.8732815; this model would be wrong about 87.3% of the time if it just always predicted that the offender was not a juvenile(s). Cohen's Kappa is a statistic that indicates how well a classifier performed as compared to how well it would have performed simply by chance. Cohen's Kappa was 0.4182 for this model. The interpretation of kappa is arbitrary because many factors can influence the magnitude of kappa, including number of observations in the test data. Additionally, there is no standard agreement for what constitutes "good" or "bad" kappa values, but Landis and Koch (1977) characterize kappa values 0.80 to 1.0 as excellent, 0.60 to 0.79 as good, 0.40 to 0.59 as moderate, 0.20 to 0.39 as fair, and less than 0.20 as poor. In this case, the model had a kappa value of 0.4182, and according to Landis and Koch (1977) guidelines, the model performs moderately well. Finally, the f-score is a weighted average of the true positive rate and precision, and for this model, was 0.4745254. An f-score of 1.0 indicates perfect precision and true positive rates, so this model could be said to perform moderately well according to the f-score. However, in machine learning, Cohen's Kappa is preferred to assess model performance because the f-score does not take true negatives into account (Powers, 2015).

XGBRF model 2 performed relatively well, but the lambda parameter was at the high end of the search range, meaning that it is likely there is a better fit value for lambda, and another parameter search was performed to determine if a better model could be attained. Table 1.3 below shows the parameter ranges searched for XGBRF model 3, and the optimized parameters returned from that search that were used in the predictive model.

Table 2.4 Parameter Search for Random Forest Model 3

Parameter Name	Parameter Range Searched	Parameter Value Returned
Nrounds	20 to 40	32
Max Depth	20 to 35	22
Eta	.35, .4, .45, .5, .55, .6	0.4
Lambda	3, 3.1, 3.5, 3.7, 4, 4.1, 4.2	3.5

For model 3, the optimal number of trees is 32, the maximum depth of trees is 22, and the space between iterations is 0.4.

XGB Random Forest Model 3 Performance

The importance matrix plot for model 3 is displayed below in figure 1.2. For this model, the five most important variables are whether the incident was a felony or misdemeanor, followed by whether the incident occurred in 2018, whether the suspect was male or female, whether the arrest was "on view", and whether the suspect was a resident of Fairfax County. Whether the incident was a theft, or whether the arrest occurred near a school, community center, or leisure location were in the top ten most important variables. These variables are the most important to the model, or contribute the most to the model, in terms of making a prediction.

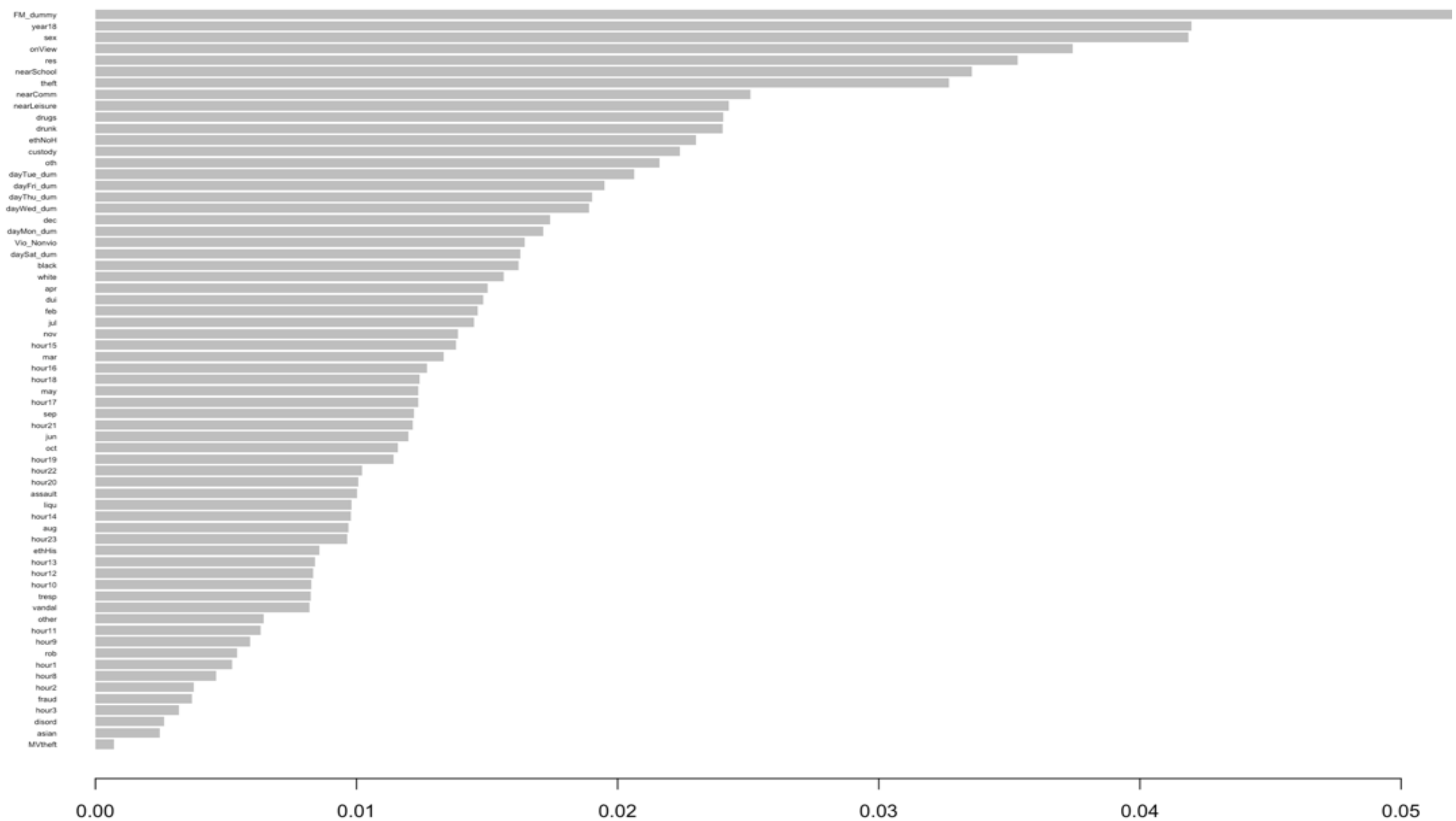

Figure 1.2 Plot of Importance Matrix for XGB Random Forest Model 3

Table 2.5 Confusion Matrix for XGBRF Model 3

	Actual “no”	Actual “yes”
Predicted “no”	46580	4308
Predicted “yes”	1633	2688

The accuracy of XGBRF model 3 is 0.8972812, or about 90%, indicating this model makes correct classifications about 90% of the time. The error rate for this model is 0.1027188, indicating this model is wrong about 10.3% of the time. The true positive rate is 0.3842196; this model predicts “yes” about 38.4% of the time that the correct classification is “yes.” Of all the observed “yes” values in the data, this model predicted 38.4% of them correctly. The false positive rate is 0.03368191; this model predicts “yes” when the correct classification is “no” about 3.4% of the time. The true negative rate for this model is 0.9663181; when the correct classification is “no,” this model predicts “no” about 96.6% of the time. Of all the “no” observed values in the dataset, this model predicted 96.6% of them correctly. The precision for this model is 0.6220782, indicating that when the model predicts “yes,” it is correct about 62% of the time. The prevalence remains the same as this is the same test data, at 0.1267185, or 12.7% of the sample is made of up incidents where a juvenile was arrested.

The null error rate for this model is 0.878172. The model would be wrong about 88% of the time if it always predicted the majority class, “no.” Cohen’s Kappa for this model is 0.4188, a moderate performance according to the Landis and Koch (1977) interpretation guidelines. The f-score for this model is 0.4750376 and based on this statistic the model could be said to perform moderately well. However, the issues discussed above with f-scores and machine learning apply, and this statistic should be interpreted carefully, or Cohen's Kappa should be used to better describe the performance of this model.

XGBRF model 3 performed relatively well, but for thoroughness, another parameter search was performed and table 2.4 below shows the parameters searched and parameters returned for XGBRF model 4.

Table 2.6 Parameter Search for Random Forest Model 4

Parameter Name	Parameter Range Searched	Parameter Value Returned
Nrounds	20 to 40	32
Max Depth	15 to 30	23
Eta	0.3, .35, .4, .45, .5, .55, .6, .65	0.3
Lambda	2, 2.3, 2.6, 2.9, 3, 3.1, 3.5, 3.7, 3.9, 4, 4.1	3.9

According to the parameter search for model 4, the optimal number of trees is 32, the maximum depth of trees is 23, the space between iterations or learning rate should be

0.3. *Lambda* is larger for this model, meaning some coefficients are pushed closer to zero.

XGB Random Forest Model 4 Performance

The importance matrix plot for this model is displayed below in figure 1.3. For this model, the five most important variables are whether the incident was a felony or misdemeanor, followed by whether the suspect was male or female, whether the suspect was a resident of Fairfax County, whether the incident occurred in 2018, and whether the arrest was "onview." Whether the arrest was for a theft, or near a school, community center, or leisure location were within the top ten most contributive variables.

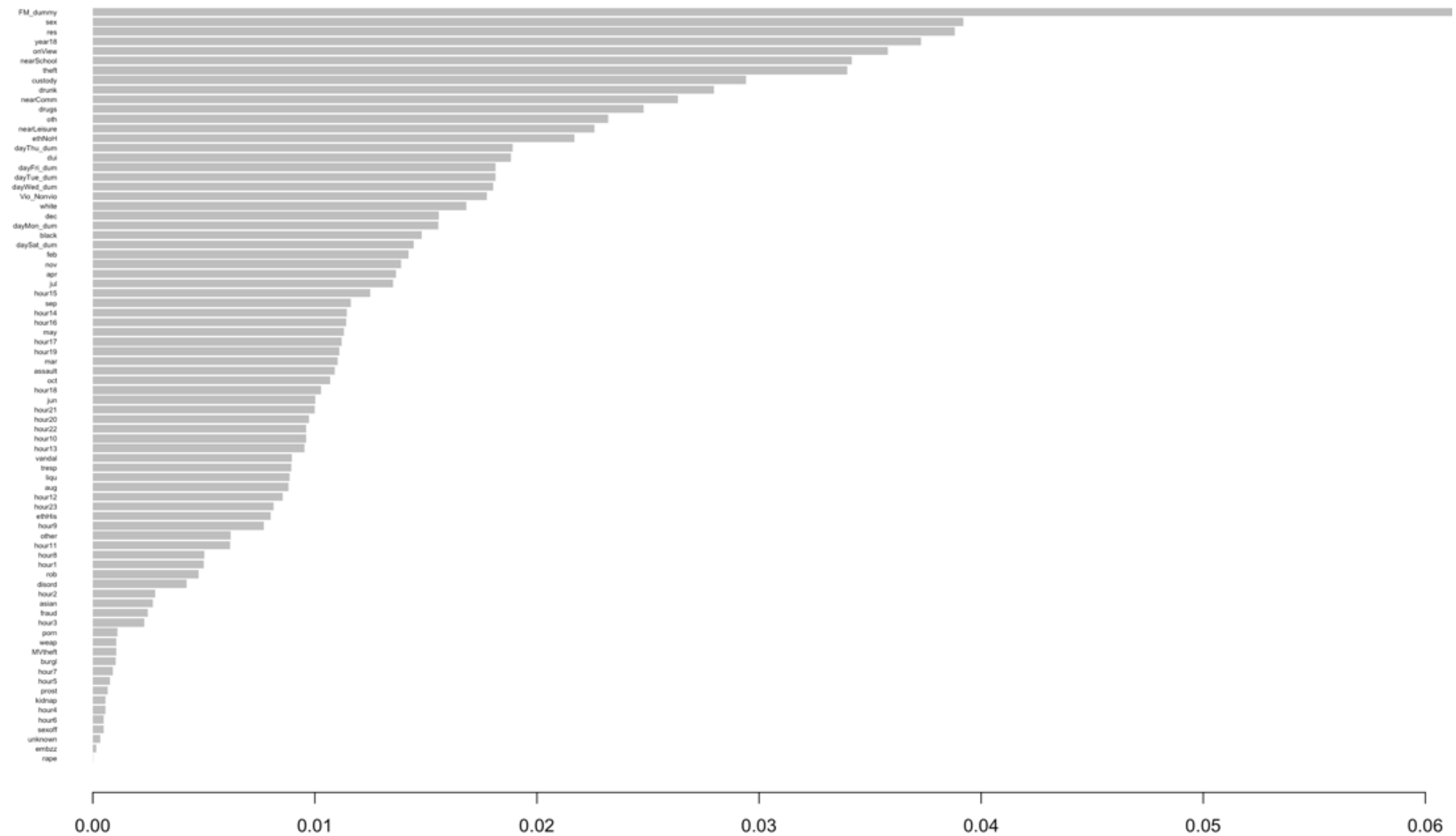

Figure 1.3 Plot of Importance Matrix for XGB Random Forest Model 4

Table 2.7 Confusion Matrix for XGBRF Model 4

	Actual "no"	Actual "yes"
Predicted "no"	46847	4504
Predicted "yes"	1366	2492

The accuracy of XGBRF model 4 is 0.8936768, or about 89%, indicating this model makes correct predictions about 89.4% of the time. The error rate for this model is 0.1063232, indicating this model is wrong about 10.6% of the time. The true positive rate is 0.3562035; this model predicts "yes" about 35.6% of the time that the correct classification is "yes." Of all the observed "yes" values in the dataset, this model predicted 35.6% of them correctly. The false positive rate is 0.02833261; this model predicts "yes" when the correct classification is "no" about 2.8% of the time. The true negative rate for this model is 0.9716674; when the correct classification is "no," this model predicts "no" about 97.1% of the time. The precision for this model is 0.6459305, indicating that when the model predicts "yes," it is correct about 64.6% of the time. The prevalence remains the same as this is the same test data, at 0.1267185, or 12.7% of the sample is made of up incidents where a juvenile was arrested.

The null error rate for this model is 0.8732815. The model would be wrong about 87% of the time if it always predicted the majority class, "no." Cohen's Kappa for this model is 0.4056, a moderate performance according to the Landis and Koch (1977) interpretation guidelines. The f-score for this model is 0.4591855 and based on this statistic the model could be said to perform moderately well. However, the issues discussed above with f-scores and machine learning apply, and this statistic should be interpreted carefully, because it does not consider true negatives, or Cohen's Kappa should be used to better describe the performance of this model.

For the final XGBRF model, some of the variables that consistently under-performed in the importance matrices were eliminated to determine if this would improve model performance. These include whether the incident was a sex offense, whether the incident was an arrest for pornography, whether the incident was embezzlement, whether the incident was possession/sale/receipt of stolen property (FCPD categorizes the "stolen property" offenses under one category name), whether the offense was a rape, prostitution, burglary, kidnapping, or weapons offense, and whether the incident occurred at 4:00, 5:00, 6:00, or 7:00 a.m. Table 2.5 below displays the parameter ranges searched for this model, and the tuned parameters returned from that search. The parameter ranges for this model were kept the same as for model 4, to determine if eliminating the un-important variables would influence the parameters returned and the overall performance of the model.

Table 2.8 Parameter Search for Random Forest Model 5

Parameter Name	Parameter Range Searched	Parameter Value Returned
Nrounds	20 to 40	33
Max Depth	15 to 30	26
Eta	0.3, .35, .4, .45, .5, .55, .6, .65	0.35

Lambda	2, 2.3, 2.6, 2.9, 3, 3.1, 3.5, 3.7, 3.9, 4, 4.1	3.7

Eliminating the unimportant variables did influence the parameters that were returned in the parameter search. The parameter values for *max depth* and *lambda* decreased from model 4, and *nrounds* and *eta* increased marginally. According to this search the ideal number of trees is 33, the maximum depth of the trees should be 26, the size of the deviation between iterations (model learning rate) should be 0.35, and the amount coefficients are reduced to avoid overfitting should be 3.7. XGBRF model 5 was fit using these parameters, and the test data excluding the eliminated unimportant variables.

XGB Random Forest Model 5 Performance

The importance matrix for XGBRF model 5 is displayed below in figure 1.4. For this model, the five most important variables returned are whether the incident was a felony or misdemeanor, whether it occurred in 2018, whether the suspect was a resident of the county, whether the suspect was male or female, and whether the arrest was on view. The arrest being for theft, or being near a school, community center, or leisure location were also in the top ten most contributive variables. These variables have remained the most important factors in making predictions in all five random forest models.

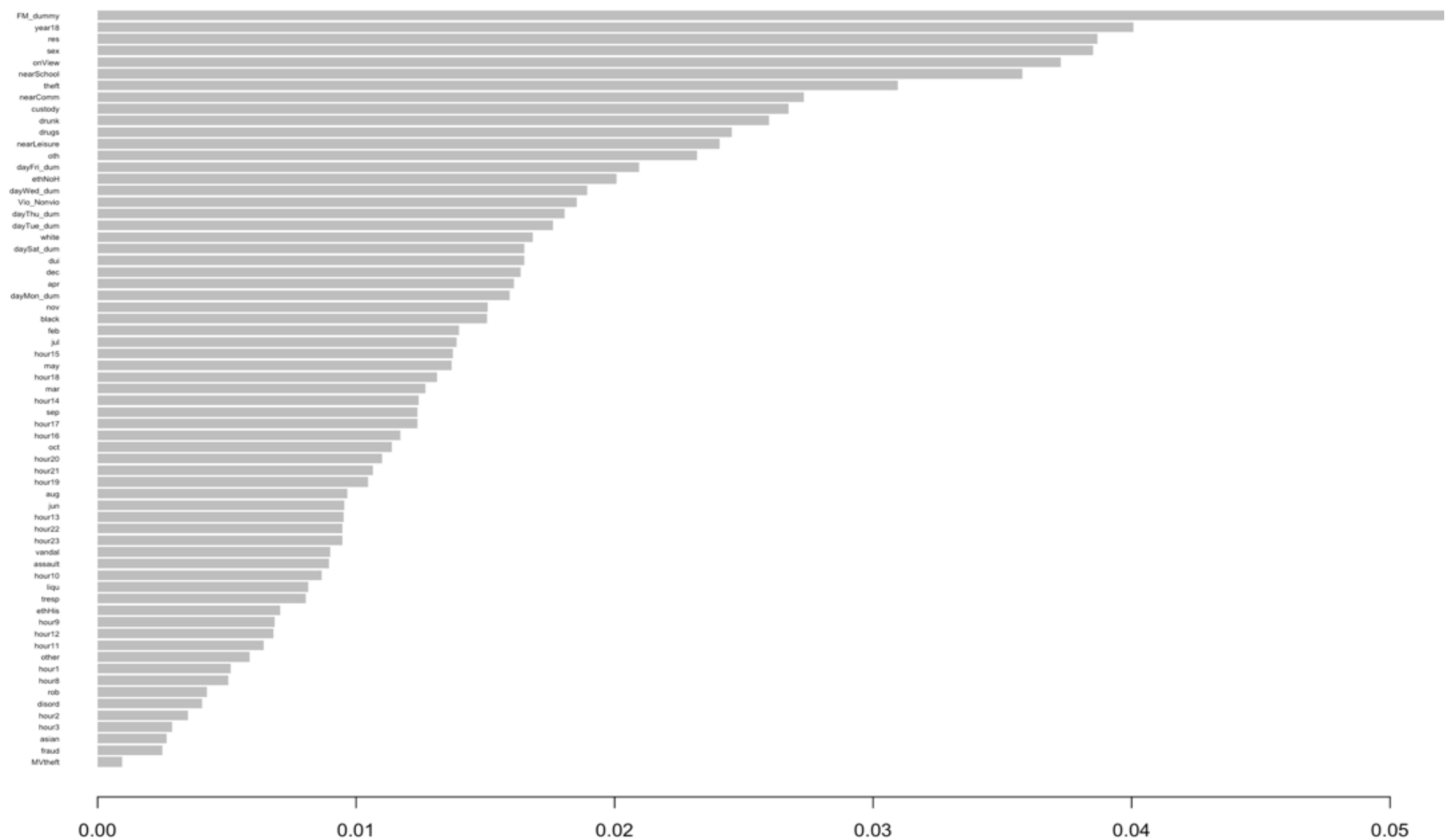

Figure 1.4 Plot of Importance Matrix for XGB Random Forest Model 5

Table 2.9 Confusion Matrix for XGBRF Model 5

	Actual “no”	Actual “yes”
Predicted “no”	46806	4467
Predicted “yes”	1407	2529

The accuracy of this model is 0.8936043, indicating that overall, the model is correct about 89% of the time. The error rate for this model is 0.1063957, indicating the model is wrong about 10.6% of the time. The true positive rate is 0.3614923, meaning this model predicts “yes” when the correct value is “yes” about 36% of the time. Of all the observed “yes” values in the dataset, this model predicted 36% of them correctly. The false positive rate for this model is 0.029183; this model predicts “no” when the correct value is “yes” about 2.9% of the time. The true negative rate of this model is 0.970817; this model correctly predicts “no” about 97% of the time. The precision for this model is 0.6425305; when this model predicts “yes,” it is correct about 64% of the time. The prevalence remains the same at 0.1267185 as this is the same sample of arrests with 12.7% being arrests of juveniles.

The null error rate for this model is 0.8732815. This model would be incorrect 87.3% of the time if it always predicted “no.” The Cohen’s Kappa for this model is 0.4087, and this indicates that it performs moderately well when compared to a model that would make these predictions by random chance. The F score for this model is 0.4626784, indicating it performs moderately well, but as mentioned above, f-scores used for machine learning binary classifiers should be interpreted with caution, and Kappa should be used where possible.

Comparing XGB Random Forest Models 2-5

Table 3.1 below displays all of the model confusion matrix values and null error rates, kappas, and f-scores for ease of comparison.

Table 3.1 All XGBRF Model Confusion Matrix Results

Confusion Matrix Rates	Model 2	Model 3	Model 4	Model 5
Accuracy	0.8922096	0.8972812	0.8936768	0.8936043
Error Rate	0.1077904	0.1027188	0.1063232	0.1063957
True Positive	0.3840766	0.3842196	0.3562035	0.3614923
False Positive	0.0340572	0. 3368191	0.02833261	0.029183
True Negative	0.966198	0.9663181	0.9716674	0.970817
Precision	0.6206976	0.6220782	0.6459305	0.6425305
Prevalence	0.1267185	0.1267185	0.1267185	0.1267185
Null Error	0.8732815	0.878172	0.8732815	0.8732815
Cohen's Kappa	0.4182	0.4188	0.4056	0.4087
F-Score	0.4745254	0.4750376	0.4591855	0.4626784

Model 2 had an accuracy of 89.2%, while model 5 had an accuracy of 89.3%. The model with the highest accuracy was model 2, with 89.7%. The true positive rate, however, decreased from model 2 to model 5, from 38.4% to 36.1%, while the true negative rate increased marginally. The false positive rate also changed from model 2 to model 5, from 3.4% to 2.9%. The model precision increased from model 2 to model 5, from 62% to 64%. The null error rate stayed almost the same from models 2 to 5.

Cohen's Kappa also stayed within the same range for models 2 to 5, all returning a value indicating "moderate/fair" performance (Landis and Koch, 1977). The f-scores for the models all stayed virtually the same as well and indicate fair to moderate performance, but again these scores should be interpreted with caution for binary classifiers.

LASSO (Least Absolute Shrinkage and Selection Operator) Logistic Regression Models

The focal point of this dissertation is to create and test predictive models to determine if juvenile age status can be predicted using characteristics of an incident. To explore this question, this dissertation uses two different predictive models with the goal of determining accuracy in predicting juvenile age status as well as exploring which types of models perform the most effectively. The second type of predictive model is LASSO logistic regression, a regularized model that can make predictions about a binary outcome (juvenile Y/N). Using the same data split into 20/80 portions randomly using R (R Core Team, 2020), the numeric binary datasets were used to perform k-fold cross validation for lasso regression, using 10 folds as is standard practice.

Figure 2.1 below is a plot of the 10-fold cross-validated mean squared error for the lasso model. The top numbering of the plot indicates the number of variables the model is using, going from all variables (top left corner) to more conservative models (top right corner). This function helps the optimization of lasso in terms of choosing the best value of *lambda*.

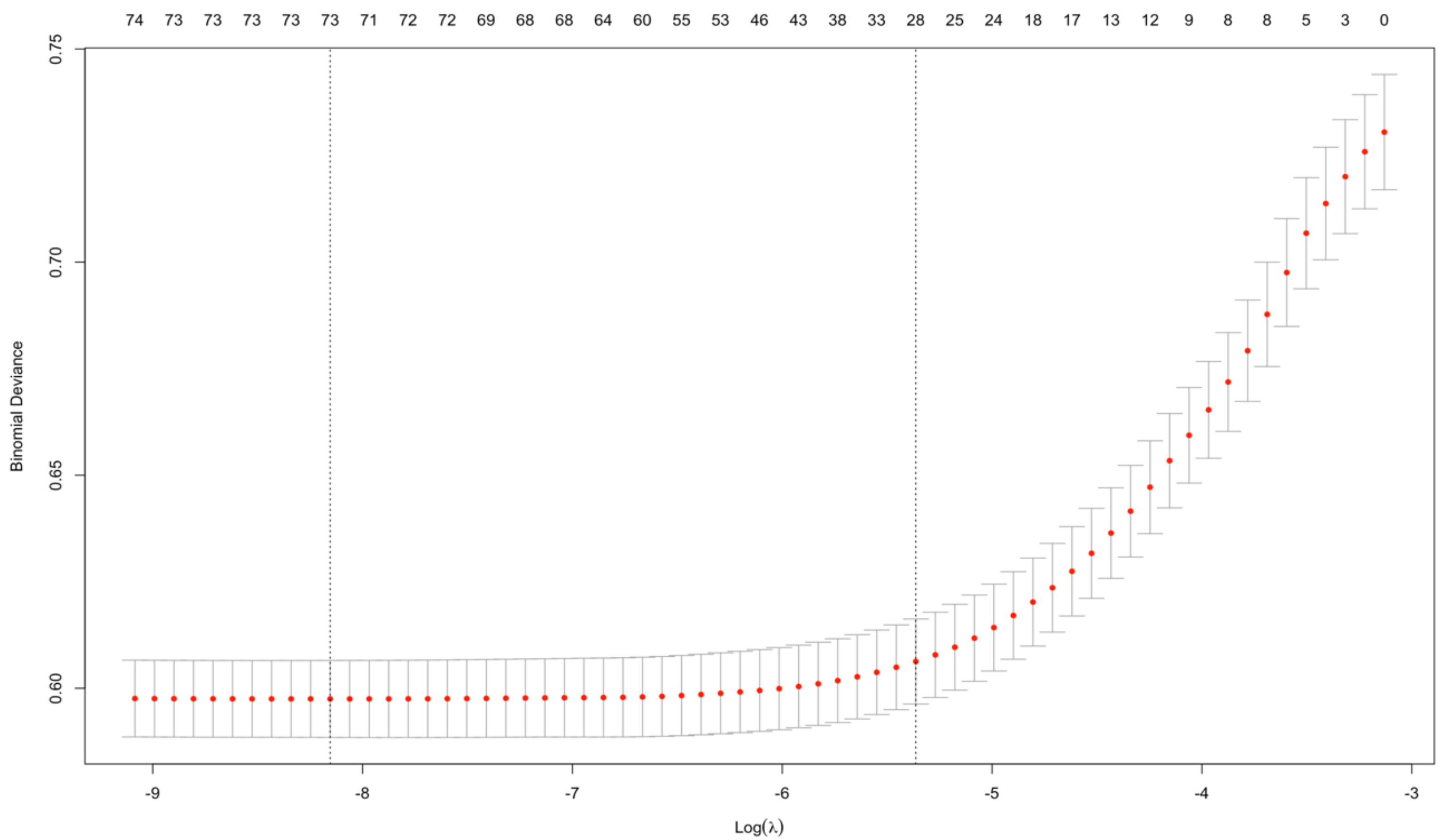

Figure 2.1 Plot of Cross-Validated MSE for LASSO Model

As the number of variables used in the model decreases, the log of lambda also decreases. After performing the cross validation on the training data, *lambda.min* was returned as a value of 0.0003154386. *Lambda.1se* was returned as a value of 0.004684163. These two lambda values are used to fit two different lasso models to make predictions for the testing data. Each model produced a column in the dataset containing probabilities that the arrestee was juvenile. These numeric probabilities were converted to binary yes/no values if they were greater than or less than 0.5. The two lasso models are compared in terms of how well they were able to predict juvenile age status of arrestees. Model 1 uses *lambda.min* as the parameter, and model 2 uses *lambda.1se* as the parameter. *Lambda.min* is the optimal lambda value for this data, and *lambda.1se* is the largest value of *lambda* where error is still within one standard deviation of the cross validated errors for *lambda.min*. The latter is a more conservative model as it accounts for some error, while a model using *lambda.min* is at risk of being overfit.

Lasso Logistic Regression Model 1 Performance

Table 3.2 Confusion Matrix for LASSO Model 1

	Actual "no"	Actual "yes"
Predicted "no"	47570	6303
Predicted "yes"	643	693

For the first lasso model using *lambda.min* (0.0003154386), the confusion matrix showed an overall accuracy of 0.8741872, or about 87.4%. The error rate for lasso model 1 is 0.1258128; this model is incorrect about 12.5% of the time. The true positive for this model is 0.0990566; this model correctly predicts "yes" about 9.9% of the time. Of all the observed "yes" values in this dataset, this model only correctly predicted 9.9% of them. The false positive rate for this model is 0.01333665; when the correct classification is "no," this model predicts "yes" about 1.3% of the time. The true negative rate for this model is 0.9866634; the model correctly predicts "no" about 98.6% of the time. The model precision is 0.5187126; when this model predicts "yes," it is correct about 51.8% of the time. The prevalence, or how often "yes" actually occurs in the data is 0.1267185, or 12.7%.

The null error rate for lasso model 1 is 0.8732815; this model would be wrong about 87.3% of the time if it always predicted "no." Cohen's Kappa for this model is 0.131, indicating poor model performance (Landis and Koch, 1977). The f-score for this model is 0.1663466, and though f-scores should be interpreted with caution in machine learning applications, this score also reflects "poor" performance as shown with kappa.

Lasso Logistic Regression Model 2 Performance

Table 3.3 Confusion Matrix for LASSO Model 2

	Actual "no"	Actual "yes"
Predicted "no"	48084	6877
Predicted "yes"	129	119

For the second lasso model, *lambda* was set to the *lambda.1se* value, 0.004684163. This is the largest value of *lambda* where the error is still within one standard deviation of the cross validated errors for *lambda.min*. This lambda is used for purposes of being conservative and taking into account the normal error that occurs in regression analyses. The accuracy of lasso model 2 using *lambda.1se* is 0.8731004, or about 87.3%. The error rate is 0.1268996, indicating this model makes incorrect predictions about 12.7% of the time. The true positive rate is 0.01700972; this model correctly classifies "yes" about 1.7% of the time. Of all the observed "yes" values in the data, this model only predicted 1.7% of them correctly. The false positive rate is 0.002675627; this model predicts "yes" incorrectly about .26% of the time. The true negative rate for this model is 0.9973244; this model correctly predicts "no" about 99.7% of the time. The precision of this model is 0.4798387; when this model predicts "yes," it is correct about 47.9% of the time. The prevalence remains 0.1267185, or 12.7%.

The null error rate for this model is 0.8732815; if this model always predicted "no," it would be wrong about 87.3% of the time. Cohen's Kappa for this model is 0.0244, indicating poor performance (Landis and Koch, 1977). The f-score also indicates poor performance, at 0.03285478. In the table below, the lasso logistic regression results are summarized. Overall, the lasso models, whether using *lambda.min* or *lambda.1se*, performed poorly compared to the XGB random forest models.

Comparing Lasso Logistic Regression Models 1-2

Table 3.2 below displays the results of the confusion matrices for lasso models 1 and 2. These models performed markedly less well than the random forest models, and the comparisons are detailed in the following sections.

Table 3.4 All LASSO Logistic Regression Confusion Matrix Results

Confusion Matrix Rates	Model 1	Model 2
Accuracy	0.8741872	0.8731004
Error Rate	0.1258128	0.1268996
True Positive	0.0990566	0.01700972
False Positive	0.01333665	0.002675627
True Negative	0.9866634	0.9973244
Precision	0.5187126	0.4798387
Prevalence	0.1267185	0.1267185
Null Error	0.8732815	0.8732815
Cohen's Kappa	0.131	0.0244
F-Score	0.1663466	0.03285478

For the lasso models, there is no improvement from model 1 using *lambda.min*, and model two using *lambda.1se*. Accuracy decreased from 87.4% to about 87.3%, and error rate increased from about 12.5% to 12.6%. The true positive rate also decreased between models, from 9.9% to 1.7%. The false positive rate decreased from lasso model 1 to lasso model 2, from 1.3% to about 0.26%. The true negative rate increased marginally from model 1 to model 2, from 98.6% to 99.7%. Model precision decreased from model 1 to model 2, from 51.8% to 47.9%. The null error rate remained the same at 87.3% in both model 1 and model 2. Both the Cohen's Kappa and the f-score values fall within the same range for both models; the generally accepted guidelines (Landis and Koch, 1977) for interpreting these values shows that both models performed poorly.

XGB Random Forest Models as Compared to Lasso Logistic Regression Models

Overall, the XGB random forest models performed better than the lasso models according to the confusion matrices for each. Though comparing the two different models in a statistical fashion is not practical, the confusion matrices show how well each model generally performs, and assertions can be made about which model performed the best. In the below table, all the model confusion matrices are summarized alongside one another, for both the random forest and lasso models.

Table 3.5 All Confusion Matrix Results for Random Forest and LASSO Models

	XGB Random Forest Models				LASSO Models	
Confusion Matrix Rates	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2
Accuracy	0.8922096	0.8972812	0.8936768	0.8936043	0.8741872	0.8731004
Error Rate	0.1077904	0.1027188	0.1063232	0.1063957	0.1258128	0.1268996
True Positive	0.3840766	0.3842196	0.3562035	0.3614923	0.0990566	0.01700972
False Positive	0.0340572	0.03368191	0.02833261	0.029183	0.01333665	0.002675627
True Negative	0.966198	0.9663181	0.9716674	0.970817	0.9866634	0.9973244
Precision	0.6206976	0.6220782	0.6459305	0.6425305	0.5187126	0.4798387
Prevalence	0.1267185	0.1267185	0.1267185	0.1267185	0.1267185	0.1267185
Null Error	0.8732815	0.878172	0.8732815	0.8732815	0.8732815	0.8732815
Cohen’s Kappa	0.4182	0.4188	0.4056	0.4087	0.131	0.0244
F-Score	0.4745254	0.4750376	0.4591855	0.4626784	0.1663466	0.03285478

There are two reasons why the lasso models likely did not perform as well as the random forest models. First, lasso logistic regression is a simple additive algorithm; the model assumes that every variable makes a unique contribution to the model, while XGboosted decision trees are interactive. XGboosted trees allow for multiplicative relationships between independent and dependent variables; one of the advantages to XGboosted random forests is the ability to explore interaction effects between the variables. For lasso to be able to capture these kinds of complicated relationships, interaction effects would need to be specified manually. Regarding juvenile crime, there are many interactions and relationships between variables that are non-linear, for example, time of day, day of week, month of year, and location. Juveniles are doing different routine activities at different times of the day on different days of the week, during different months of the year, at different locations, which can impact the likelihood that a juvenile is committing a crime during a given time, day, and month, and at a given place (Clarke, 1995; Cohen and Felson, 1979, Weisburd, 2009).

Additionally, lasso models are prone to over-fitting with many input features, of which this data had 93 once all the variables were converted to binary format. The lasso models likely performed well on the training data, but when presented with the test data, they likely underperformed due to overfitting because there were so many input variables. For the random forest models, using XGboost to increase the aggregate complexity of many smaller trees (each tree tries to correct the prediction errors of the one before it, creating many small trees that have a strong prediction ability when they are working together) and setting a limit on nrounds (number of trees) and max depth (number of branches), are effective ways to avoid overfitting. Unconstrained random forest models (no limit set for nrounds or max depth) are very prone to over-fitting because they would just “memorize” the training data entirely, but then perform poorly when presented with the new testing data. However, though the XGboosted random forest models performed much better than the lasso models, there are still limitations to consider, discussed further below.

Overall, the XGboosted random forest model 2 was the best fit model in terms of accuracy and the other confusion matrix values, but it is still not a perfect model. The accuracy of model 2 is 89.7%, which is relatively good, leaving only a 10.2% error rate. However, this is evidence that the relationships and variables that indicate whether a crime was committed by a juvenile are complex, and a lot more time, data, computing power, and model tuning (more than possible for the scope of this dissertation) are needed to create a more accurate model.

CHAPTER 5: DISCUSSION AND LIMITATIONS

This dissertation set out to test whether machine learning models could be used to predict the juvenile age status of an offender using characteristics of the offense. The purpose of this exercise is to better understand the "dark figure of juvenile crime"—informed by criminological theory—so that more targeted and specific prevention measures for crime patterns created by youth can be more effectively addressed. Using arrest data from Fairfax County, Virginia, four XGboosted random forest models and two lasso logistic regression models were developed, tuned, tested, and evaluated to determine their accuracy in predicting whether a crime (in this case, a crime that results in an arrest) is committed by a juvenile (under 18). The XGboosted random forest models outperformed the lasso logistic regression models, with random forest model 2 having an accuracy rate of 89.7%, and a kappa value of 0.4188, indicating moderate to fair model performance (Landis and Koch, 1977). Plainly, random forest model 2 was moderately better at predicting juvenile status than a model that would make the same predictions just by chance; evidence that the model is having some success at making accurate predictions by using the variables in the data and the set parameters in a meaningful way, rather than just being accurate by "throw and stick" happenstance.

While the random forest models performed moderately well, the lasso logistic regression models had accuracies of about 87%, and had kappa values of less than 0.2, indicating very poor model performance (Landis and Koch, 1977). In other words, the lasso logistic regression models were not very different than a model that would make the same predictions by random chance. The likely reason for this is that lasso models are simple additive models: they do not capture the complexities of variable relationships and interactions very well, but are able to make accurate predictions about more straightforward relationships. The random forest models, on the other hand, capture non-linear relationships and interactions well. This is due to the nature of XGboosted decision trees; each decision node in the tree depends on the last one, creating a path of variable interactions that a lasso model is not able to capture, unless specific interaction terms are set manually.

Given the non-linear nature of juvenile crime, several variables in the model were important in making the predictions. In the importance matrices, some of the most contributive (important) variables included if the incident was a felony or a misdemeanor, if the arrest was "on view," arrestee residency status, if the offense was a theft, and if the arrest was near a school, near a leisure site, or near a community center. These variables align with juvenile crime theory. Juvenile crime often consists of non-violent, low-level offending. The variable indicating if an incident was a misdemeanor was one of the most important contributors to all models, and most incidents that juveniles were arrested for were non-violent. These observations align with developmental and social learning theories, as well as opportunity and situational theories. Juvenile offending often occurs in unsupervised groups of peers, as significant social rewards are offered in return for impressing one's peers; sometimes, this involves risk taking behavior such as criminal offending (Burgess and Akers, 1996; Hirschi, 1969; Osgood et al., 1996). Further, the opportunities for this kind of group "hanging out" where offending tends to occur are concentrated in certain spatial and temporal spaces; for example, places such as a mall where juveniles are away from parents/guardians, and at times when they are not participating in their required daily routine activities (e.g, school, sports, job, etc.) (Weisburd et al., 2009; Hoeben and Weerman, 2014; Osgood et al., 1996). Finally, the importance of the felony/misdemeanor variable in predicting juvenile crime aligns with Moffit's (1993) developmental taxonomy, in which the most common kind of offenders, adolescent limited offenders, engage in low-level offending throughout a period during adolescence and then cease offending when they enter adulthood. This alignment is also supported by the offense type "theft" being one of the top predictors of juvenile crime; most juvenile offending is low-level, and adolescent limited. This model may be effective at predicting juvenile offenders when the offenses are indicative of "adolescent limited" offenders, but the model may mistake the more serious offenses committed by "life course persistent" juveniles for offenses committed by adults, and this is discussed further in the limitations section below.

Residency status was also an important contributor to the predictive model, with most juvenile arrests being of residents of the county. This aligns with routine activities theory in two ways; juveniles only have certain times outside of their required routine activities that they are unsupervised and have an opportunity to offend, and juveniles are more likely to offend in their hometown or in areas that they are very familiar with, that are a part of their routine activities or daily routine. Youth feel comfortable in these places, and their perceptions of the risk of being caught and/or apprehended are low, due to their familiarity with the location (and with the risks present there) (Brantingham and Brantingham, 1993). In addition to being more likely to offend within a close to home "comfort zone," juveniles may also be more likely to offend close to their homes, schools, or local "hang out" spots due to their lack of access to transportation, for example, if they do not have a driver's license or if they do not have a vehicle (Brantingham and Brantingham, 1993). These offending locations also align with routine activities theory; youth have a greater propensity to offend when they are on their way to or from their routine activity locations such as school, or leisure places such as a mall or arcade (Brantingham and Brantingham, 1993; Osgood et al., 1996; Roncek and Lobosco, 1983; Weisburd et al., 2009; Wilcox 1973).

Finally, "onview" arrest is an arrest data specific variable but may also have to do with biases or differences in police discretion and the offenses that are vs. are not discovered. Most juvenile offenses are low-level offenses that often go unsolved, and/or are never discovered, and if an offense is discovered after the fact, especially for low-level offenses that occur in unsupervised settings, the probability of catching the offender is usually very low. The offenses that are discovered and lead to an arrest may be more likely to be the events where juveniles are caught in the act and arrested in person, on or near the scene. For example, an officer sees a group of kids vandalizing a building and makes arrests on the scene. Crimes committed by adults tend to be more serious in nature (e.g., robbery, homicide, sex offenses) and thus lead to a police investigation and subsequently, an arrest. The low-level, non-violent offenses committed by juveniles are likely to go undiscovered or unreported and will likely not result in an arrest unless an officer catches a juvenile in the act of offending, and even then, arrest is discretionary. The kinds of offenses that juveniles are caught committing vs. those that are not caught and never discovered may be different in the nature of the offense, or in spatial or temporal contexts. This is a hypothesis that might be explored in the future with a predictive machine learning model such as this, to predict if offenses discovered after the fact might have been committed by juveniles, and to discover previously unknown patterns of juvenile offending.

The goal of conducting this exploratory analysis was to determine if it would be at all possible to develop a machine learning model that could predict the juvenile status of offenders in events where the offender (and their age) is unknown. The analysis here gets closer to that, but this dissertation focuses specifically on incidents that have resulted in arrests, because of the need for validation of the model's accuracy against data where the ages of arrestees is known. The are several practical applications of such a tool. Future development and refinement of a machine learning tool that can predict the juvenile status of suspects/offenders in offenses where the perpetrator(s) is unknown can lead to advancements in juvenile crime prevention, intervention, and evidence-based policing strategies, discussed below. In terms of scientific application and contribution to the field, machine learning tools that are able to make these predictions accurately can assist in uncovering some of the "dark figure" of juvenile crime; where offenses are unknown or unrecorded, because they were either not reported or discovered. On a smaller scale, this dissertation is an initial exploration into the possibility of using machine learning to predict the age group of offenders based on incident characteristics.

Paving the Way: The Future for Juvenile Crime Policy and Prevention

The development of a machine learning tool that can accurately predict the juvenile status of an offender in incidents where the offender is unknown and an arrest has not been made, is the long-term goal. This dissertation is a step toward that goal and illustrates that given information about an incident such as location, time of day, day of week, incident severity (felony/misdemeanor), and type of offense, random forest models can predict juvenile status with about 90% accuracy (XGBRF model 2). Predicting the juvenile status of offenders in incidents where arrests may not necessarily be made, or in events where the offender(s) are not known, could be advantageous to police departments and policymakers. In addition to uncovering some of the "dark figure" of juvenile crime, being able to accurately predict juvenile status of offenders in incidents where offenders are unknown would lead to a better understanding of where, when, and how juvenile crime occurs. Having an increased understanding of spatial and temporal patterns of juvenile crime would allow police departments to better allocate patrol to "hot spot" locations and/or times. Police agencies could develop more accurate problem-oriented policing strategies, such as focusing on certain places at certain times that they had not previously known about, given the gaps in the current data about juvenile crime. Police and community stakeholders may also be able to work together to develop preventative policy that focuses on non-arrest interventions or addressing the roots of juvenile offending.

A real-life example of this might look like the following: police in Sunnytown use this machine learning tool to predict the juvenile status of offenders in their crime and/or calls for service data where arrests have not been made, where incidents have not been cleared. Using this strategy, they discover a disproportionate number of incidents in a certain neighborhood, specifically near Cloud Street and Ray Avenue. These incidents involve relatively minor offenses, such as disorderly conduct, fighting, vandalism, or evidence of drug use/sales/paraphernalia. When these incidents are called in by the residents of Cloud Street and Ray Ave., the police often arrive to the scene too late; the perpetrators are long gone by the time they arrive, having been suspicious of being caught and leaving the area soon after they engage in the problematic behaviors. The police notice that numbers of incidents involving juveniles (as predicted by the random forest model) indicates that incidents in this area spike in the afternoon hours (3:00-5:30 p.m.) and especially on weekends. Police begin to implement routine patrols in that area, in small doses from 3:00-5:30 p.m. on weekdays and throughout the day on weekends. The police manage to speak to some of the juveniles that "hang out" in this area, and they discover that the kids are hanging out on the streets, having nothing better to do, because the nearby skatepark has been closed off; their hobby has been taken away, their normal routine activities disrupted. The police take this information and bring it up at the next town hall meeting to determine why the park had been closed, and they raise the issue that if the youth had something healthy to do (a hobby, such as skating) and somewhere specifically for them (youth, teens, young people) to "hang out", most would likely stay out of trouble and off the street. The town stakeholders explain that the skatepark was closed because too many parents complained of injuries and kids falling. The police discuss implementing a requirement for proper skating safety gear to be worn before people are allowed to enter the skate park, and the town hires a "safety monitor" for the park to observe during open hours and ensure that these rules are being followed. The town re-opens the skate park with the new safety measures, and within several weeks the police notice that their calls for service concerning people "causing trouble" on Cloud Street and Ray Avenue have noticeably decreased. The police routinely drop by the skate park to ensure the safety procedures are followed. The park is crowded with teens almost every afternoon, skating and "hanging out" after school and on weekends. Though the police patrol the skate park regularly, they do not see anything concerning or out of the ordinary; the kids are occupied engaging in their hobbies – their usual routine activities. Police and community stakeholders can work together in this way to address the root of "juvenile crime problems," rather than simply focusing on making arrests and "cracking down" on youth.

Another way this kind of information could be beneficial to police is that it can be used to develop specific crime interventions that are not arrest focused. For example, if police discover that a large number of "fight" or "assault" incidents are happening during times when juveniles would be on their way to/from school, and in the corresponding locations, the police may be able to work with the school to develop an action plan to prevent fighting or bullying; school resource officers could patrol the routes where incidents are most common, or the school may implement a youth-driven safety program, with peer "observers" that are stationed along these routes to encourage positive interactions, discourage fighting/bullying, or to at least report what happened and who was involved to the appropriate adults so that serious situations can be addressed by the school administrators, by parents or other guardians, and/or by school counselors.

With increased knowledge of where and when juvenile offending is occurring, and what type of offending behaviors that youth engage in, police and community stakeholders can work to implement measures to intervene and prevent offending behaviors, by addressing the roots of the problems, or at the very least, increasing the perceived risk of apprehension to dissuade would be offenders (Nagin, 2013). These are only a couple of examples of the vast possibilities that arise from having a tool that can accurately predict juvenile status of offenders in events where offenders are unknown and/or where arrests do not necessarily have to take place. Knowing specific places and times where youth engage in offending behaviors can, simply, allow police to focus on the right people, at the right places, at the right times, rather than use "wide-net" strategies that are often described as being presumptive and unfair by youth (Fratello et al., 2013). Having this increased view of what is going on in terms of juvenile offending can improve police-community trust and relations as well, as this can prevent the need for "wide-net" strategies that make assumptions about youth that youth often describe as feeling unfair and disrespectful (Fratello et al., 2013).

The above are only a few of numerous examples of how a machine learning model such as this could be used for practical crime prevention and intervention purposes. In addition to the practical applications of this tool, there are many opportunities to use the new information the model produces about the spatial and temporal patterns of juvenile crime; are there other juvenile crime hot spots that have not been detected in prior analyses? How (if at all) does the effectiveness of juvenile crime interventions change, given the new information from the model? Further, if this tool were to be used by police agencies nationwide, there is potential for the creation of a national reporting system for these data, similar to the National Incident Based Reporting System (NIBRS). If a national-level database of crime data with predictions about the offenders' juvenile status existed, future research on juvenile crime patterns may be able to bring to light some of the "dark figure" of juvenile crime. Insights such as these could lead to the development of more effective, efficient, and fair juvenile crime interventions and policy. A machine learning tool with this predictive ability has numerous practical and research applications. However, the limitations of such an approach are important to consider.

Research and Practical Implications and Their Limitations

The first limitation of this study is that the ability of this machine learning model to predict the juvenile status of offenders in all crimes (not only those that result in arrest) is still unknown. The use of arrest data with ages of arrestees included was necessary for the analysis of model performance; for the purpose of this dissertation, there would be no way to tell how well/not well the models performed. Creating a machine learning model that is able to accurately predict the juvenile status of offenders in all crimes (even where offenders are unknown, or an arrest is not made) is imperative to being able to use this tool to learn more about juvenile crime.

The second limitation of this study is that the data come from only one county, in one state, and thus the variables used to build this model are specific to that area, and would differ for different states, cities, and counties. This model is not generalizable to other areas; it only performs the way that it does with the Fairfax County, VA arrest data. As mentioned above, developing a practical model for use in police departments would take months or years of development, study, and model tuning, along with the budgetary concerns that come with such a large scale and long-term effort.

Third, the lasso logistic regression model is not a good fit for the Fairfax County arrest data, and that is likely because the lasso fits linear relationships and is prone to overfitting, meaning that the model works well for the training data but under-performs on the testing data. Juvenile crime has many covariates and predicting whether an offender is juvenile based on characteristics of the incident that occurred is not linear; there are many variables and interactions between those variables to be considered, which lasso models do not capture well. The lasso also needs a dataset with less variables due to its tendency to overfit when many variables are included. This is problematic for predicting whether offenders are juvenile, as there are many variables that matter in crime incidents for increasing or decreasing the likelihood that an offender is juvenile.

Fourth, this model may be better at predicting incidents involving adolescent limited offending juveniles rather than life course persistent offenders (Moffitt, 1993), who may look like adults in the prediction model. One of the most important predictor variables in all four random forest models was whether the event was a felony or misdemeanor; often the most serious or serious violent offenses are felonies, and it is likely that if a juvenile committed these crimes, they would classify as a life-course persistent offender (Moffitt, 1993). The seriousness of the incident may cause the machine learning model to predict "adult" for these rare, but existing, serious crimes committed by juveniles. This is an important limitation to consider; this may be one of the areas of the "dark figure" of juvenile crime that remains hidden. If this predictive technology is not able to discern juveniles from adults for very serious crimes, the accuracy of patterns of life-course persistent juvenile offenses may not have the same improvements that patterns of adolescent-limited juvenile offenses may see with the implementation of this predictive technology. Although most juvenile crime is of the less serious, adolescent limited type, some juveniles do engage in very serious offending behaviors; being able to accurately predict juvenile status for both non-severe and serious juvenile offenses would be a major improvement to these predictive models.

Finally, arrest data was used to construct these models, and with that, the location of arrests. However, the location of an arrest is not necessarily the location where the crime occurred. This is relevant to juvenile crime because of juveniles' highly structured routine activities (Weisburd, 2009; 2015) that put them in certain places at certain times, making location of the crime a valuable piece of information in determining whether an offender is juvenile. Unfortunately, there is not currently a system other than self-reporting or police detective work that can determine and record where exactly crimes have occurred, especially those that lack witnesses, reporting parties, or cooperative suspects/arrestees to give this information. Arrests, however, are generally a good proxy (Weisburd, 2009). Additionally, arrest data have some bias due to the discretion that police have in making the decision to arrest or not arrest someone. For juveniles, officers can, and it appears they often do decide to handle things informally or to let youth off with a warning. Research shows that in youth-police interactions, youth are rarely arrested, only 13% (Myers, 2004) to 16% (Liederbach, 2007) of the time. Some police discretionary decisions are shaped by reform policies aimed at diverting juveniles from the criminal justice system. Liederbach (2007) finds that less than half of juveniles who committed a criminal offense were arrested, and instead of arrest, police will often tell a juvenile(s) to stop what they are doing (57%), interrogate them (37%), or threaten them with a criminal charge without really arresting or charging them (20%). In keeping with community policing initiatives, police are sometimes required to divert or direct juveniles to services, such as anti-drug or anti-gang programs, in lieu of making an arrest (Bannister, Carter, and Schafer 2001; Giblin, 2002; Withrow and Bolin, 2005). Fratello et al. (2013) note that in 2010, 23 percent of all youth arrests were handled "within the department," resulting in the youth being released. In addition, about 50 percent of cases that are referred to juvenile courts are handled informally. Informal handling of cases usually includes conditions that the youth voluntarily agree to, and upon completion, the case is dismissed (Fratello et al., 2013). Incidences where youth are warned or diverted are not recorded in official arrest data, leaving out many instances of juvenile delinquency and police contact. The research limitations of the study, for the most part, concern data quality and the issue of being able to predict juvenile status of subjects in all crimes; not just those that result in arrests. However, there are numerous limitations concerning the practical implementation of this tool, as discussed below.

In terms of practicality, in its current state, this machine learning model would likely not be a good fit to be used in practical policing applications. First, this tool will not perform at the same level for different sets of variables; one of the main practicality flaws with machine learning is that data must be in a specific, standardized format. If this model were to be implemented in every police department in America right now, it would need to be trained on every single police department's unique data and variables, because they all differ. If the data system in the department ever changed, the model would have to be re-trained on that new data format because it would not perform in the same way with different variables. There currently is no required federal level standardized arrest or crime data reporting format; the systems that do exist (i.e., NIBRS) are voluntary to report to and we must consider the differences between jurisdictions that choose to report and those that do not. This limitation highlights the need for a federal (or even state) level standard crime reporting data format.

However, even with a standardized data reporting format, variables would likely differ from state to state or town to town. For example, in this evaluation "nearLeisure" was used to determine if the arrest took place near a shopping mall, movie theatre, etc. Some towns do not have hangout spots such as shopping malls or movie theatres near them. For example, New Egypt, New Jersey, is a small rural town where even the closest shopping mall is about a 30-minute drive by car. There are no buses, subways, trains, or other forms of public transportation readily available in/near the town; those require travelling to different towns and cities by car as well. However, the local youth often hang out in a park located near the local ice cream shop, a variable that is very specific to this town. Alternative “leisure” places would need to be used in a model to accurately predict offender age group here, which also then requires studying where youth in this and similar towns “hang out” and entering those places into the model instead. This creates another practicality issue, as police departments would need the time, budget, and expertise to determine how to come up with these variables. This leads to multiple different machine learning models for multiple different cities, towns, and states, that may perform in vastly different ways.

To follow, aside from the issue of lack of standardized variables, this tool would need to be converted to an interactive user-friendly computer program where a user could, for example, click on various characteristics of the incident (i.e., click “felony” or “misdemeanor”) and the model would then produce a prediction based on that information. This requires pruning the decision tree model (entering/deleting variables, trees, branches), which may affect how well it performs. There is also the issue of time, and limited police budgets, for development of these models which would take months (or years) and requires high-powered (expensive) computers to do the job. The parameter search for the random forest models in this dissertation took about two to three entire days each to finish and produce results, on a laptop with 16 (the average is 4-8) accessible logic cores all working on running the parameter search and doing nothing else. Multiple computers may be needed so that workflow does not get interrupted because the computer needs 2-12 days (or more, depending on processing power of the computer) of uninterrupted working time, which is not practical for most (if not all) police departments. Additionally, trained personnel would be needed to perform these tasks, whether that is someone internal to the department that receives training, or an outside hire that specializes in this area. The question remains whether police departments can justify the budget allocations to hire such a person, and this is especially true in smaller police departments that do not receive large budget allocations.

This tool is not intended for use in finding suspects to make arrests, but rather to identify larger patterns of juvenile crime that may have been previously unseen. Even with an 89.7% accuracy rate, there is still a relatively large margin of error (10.2%) in the predictions that the model makes. This model was also developed using arrest data, due to the necessity of an observed age group (juvenile/not juvenile) variable to test the model's predictive accuracy. But the ultimate goal of building a model such as this would be to predict whether a crime was committed by a juvenile(s), rather than predicting whether a juvenile was arrested for a crime.

Variables in most arrest data are similar to those of crime incident data (e.g., "felony/misdemeanor," "location," "type of crime," etc.), and given the relative success of random forest models in this study, it would not be inaccurate to say that it is possible to use machine learning to predict the juvenile status of offenders based on characteristics of offenses. However, an important caveat is that the purpose of this tool is to predict the juvenile age status of offenders (suspects) in order to establish more complete juvenile crime data; to shed some light on the "dark figure" of juvenile crime, where it is unknown if, when, and where youth are offending, because those offenses are not discovered and/or not recorded. The big-picture goal of a machine learning tool with these capabilities would be to inform evidence-based, more efficient, effective, and fair juvenile crime policy.

It is relatively well-established that there are some issues with using predictive technology in the criminal legal system, for example, the racial disparities in risk assessment tools (Moore and Padavic, 2011). These kinds of errors can be harmful to minority populations and having a machine learning tool that over or under predicts juvenile offending can lead to over or under policing of juveniles, the avoidance of and reduction of which is one of the reasons why these models would be implemented in the first place. To avoid this kind of bias, the machine learning models would have to be trained and tuned most precisely to ensure that they are as accurate as possible, a process that can take considerable time, computing power, and financial means.

To sum up, machine learning has the potential to be a powerful tool when used in this way, but it comes with limitations that are important to consider. Importantly, the ability of this tool to predict the juvenile status of subjects in all crimes (not only those that result in arrests) is still unknown. It is also necessary to consider the limitations of practical implementation of this tool as well. Machine learning models can be expensive, and time consuming to develop and implement. These are essential limitations to consider when expanding the research base in this area.

Implications for Future Research

This dissertation is the first to attempt to use machine learning to make predictions about the age of offenders based on characteristics of the offense. This study serves as a jumping off point for future studies seeking to use machine learning in practical applications. There is much more research needed to determine how to create a model with better accuracy in predicting whether an offender is juvenile. First and foremost, the predictive ability of these models has not yet been tested on all crime data – where subjects may be unknown, and incidents may not ever result in arrests. These predictions can be used to compare crimes predicted to be juvenile-perpetrated, and those not to identify possible new patterns in juvenile crime. For the purposes of this dissertation, it was necessary to use arrest data so that model performance could be evaluated accurately. Future studies could also attempt to use different machine learning models to predict age ranges, rather than binary age group status, for offenders based on characteristics of the offense.

In terms of research that would further the practical application of machine learning, studies of which variables are the most important in making predictions, in which places, need to take place for more accurate models to be built. On a smaller scale, cities and towns may need to conduct studies in conjunction with their police departments to determine what some of the variables would look like specifically for those locations, i.e., for a city with no shopping malls or "typical" youth hang out spots. This dissertation opens the door to numerous possible future studies as well as the possibility of using machine learning to identify where and/or how police could more effectively prevent or intervene to reduce crime in their communities.

The possibilities are numerous concerning the application of machine learning algorithms, specifically random forest models. Random forest models like the ones in this dissertation could be used to predict the age groups (juvenile/non-juvenile) of offenders in incidents where the offender(s) is not known. This would provide valuable insight into the locations of juvenile crime, “hot spots” of crimes that were going undiscovered or unreported, and quite possibly updated juvenile crime rates/proportions for cities and towns. These advances in knowledge can lead to development of evidence-based juvenile crime policy that is more fair, more effective, and more efficient.

On a larger scale, this application may look like a computer program, or a website, that crime analysts could upload crime incident data to, then train, tune, and execute the random forest models. A more user-friendly, smaller, faster version of this may be a decision-tree point and click computer program, that takes the variables included in an uploaded dataset, converts them to binary numeric variables, and displays a decision tree path where each variable is a yes/no question about the incident, and the user can click a “yes” or “no” button, then the model brings up the corresponding next question, and so on, until a prediction is generated.

To expand this idea, an algorithm generator program could be developed, to produce a random forest model based on an uploaded police dataset, with automated parameter searching and tuning based on parameter ranges (*nrounds, max depth, eta, lambda*) that are adjusted based on conditions of the results, e.g., a function that determines from the returned parameters whether the ranges need to be expanded and searched again. For example, a function with rules set so that if *eta* is within one standard deviation of the upper or lower extreme in the range, the parameter search is expanded by *x* in the corresponding direction and the parameter search is performed again with the updated range values. Once this searching and tuning is completed, the optimal parameters would be set and stored in a machine learning model, customized to the agency/organization that uploaded the data and performed this tuning/training. However, limitations concerning time, funding, training of data analysts/detectives/etc. to use this program, computing power, and data storage space reign in this possibility for the immediate future. This is a conceivable glimpse into the distant future of this technology, far beyond the scope of this dissertation, but with advances in cloud computing and data storage, is not far beyond the realm of possibility.

Conclusion

This dissertation explores the potential for machine learning to be used as a tool in juvenile crime intervention and policy development. The aim of this dissertation was to build a predictive model that, using characteristics of an incident in question, could accurately predict the juvenile status of arrestees for known crimes. With a tool like this, that is accurately able to predict juvenile status of offenders in all crimes, evidence based policing strategies and juvenile crime policy could be improved. These models could bring to light more of the "dark figure" of juvenile crime, and with that, lead to improved policy and more accurately targeted policing interventions and strategies. These methods have the potential to greatly advance knowledge and research concerning juvenile crime. Though it is not realistic to expect that juvenile crime will cease, police-community relations concerning juveniles and juvenile outcomes from the justice system could be much improved. At the individual, personal level, a machine learning tool with this kind of predictive capability could lead to the development of evidence-based non-arrest focused juvenile crime interventions, leading to better outcomes (diversion programs, educational resources, social services, etc.) for juveniles who otherwise may have served time in a detention facility or faced other serious legal consequences that could affect the rest of their lives. Criminal-legal outcomes for countless youths could be vastly improved with a tool like machine learning that allows for a better understanding of juvenile crime patterns.

On a wider scale, if spatial and temporal patterns of juvenile offending are better understood, more effective measures can be taken to intervene and prevent juvenile offending. With these new insights into juvenile offending patterns and the resulting evidence-based policies, juvenile crime prevention and intervention could be more effective, more efficient, and more just. This machine learning tool and the information it could yield would help to improve police-community relations between youths and the police due to more accurately targeted enforcement and patrol strategies rather than wide-net ones that seem to target the wrong people, at the wrong places, at the wrong times. Further, with more effective and efficient policies in place, police officers may feel less over-burdened at work, and police agencies may incur significant cost and resource savings, that they could allocate to other areas where those resources are needed; potentially allowing law enforcement agencies to be more effective and efficient overall.

Could machine learning models be used to predict whether an offender is juvenile for crimes where the offender(s) are unknown? It is possible. The best fitting model in this dissertation had an accuracy of just about 90%; however, machine learning models will likely never be perfect. Human beings are in control of these computer models, of the functions that control them, and computers cannot figure out which questions to ask, how to ask them, which variables to include, etc., without a human programming the computer and telling it what to do. This leaves machine learning open to inevitable flaws, because no study conducted by humans can ever be perfect; humans are inherently flawed whether it be by bias, mathematical mistakes, misunderstanding, lack of knowledge, or any of the other numerous and unavoidable human faults. By extension, the machines that we create using our knowledge are also inherently flawed; there may not be a way to achieve a perfect prediction machine. The question then becomes, how accurate is enough for practical use? To this question, the answer remains unclear.

However, this does not eliminate the possibility for future use of machine learning in practical policing applications, such as identifying previously unknown hot spots of juvenile crime. The exploration of machine learning for policing applications is an endeavor that could lead to evidence-based juvenile crime policy that is more effective, fair, and efficient; a contribution to the criminal legal field worthy of the time and effort that developing such programs requires. This dissertation has demonstrated that it is possible for a machine learning model to predict whether an offender is juvenile based on characteristics of the offense but developing the technology into a practical application would take exponential amounts of money, time, and computing power, as well as the cooperation of law enforcement agencies in either individually uploading their data to create a custom model, or creating a standardized crime data reporting format for a national-level standardized model. Using machine learning to predict ages of offenders and uncover the "dark figure" of juvenile crime is certainly possible, but not yet practical.

This dissertation is a first step, a jumping off point for the future of the use of machine learning for predictive and proactive policing application.

REFERENCES

Agnew, R. (2003). An integrated theory of the adolescent peak in offending. *Youth & Society*, *34*(3), 263–299. https://doi.org/10.1177/0044118X02250094

Agnew, R. (1997). Stability and change in crime over the life course: A strain theory explanation. In T. P. Thornberry (Ed.), *Developmental theories of crime and delinquency* (pp. 101–132). New Brunswick, NJ: Transaction.

Akers, R. L. and Gang, L. (1999). Age, social learning, and social bonding in adolescent substance use. *Deviant Behavior.* 19:1–25.

Anderson, A. L., & Hughes, L. A. (2009). Exposure to situations conducive to delinquent behavior: The effects of time use, income, and transportation. *Journal of Research in Crime and Delinquency*, *46*, 5-34.

Bannister, A. J., Carter, D. L., & Schafer, J. (2001). A national police survey on juvenile curfews. *Journal of Criminal Justice*, 29, 233−240.

Baumgartner, M. P. (1988). *The moral order of a suburb*. New York: Oxford University Press.

Bernasco W. and Block R. (2011). Robberies in Chicago: A block-level analysis of the influence of crime generators, crime attractors, and offender anchor points. *Journal of Research in Crime and Delinquency.* 48(1):33–57

Bichler, G., Malm, A., & Enriquez, J. (2014). Magnetic facilities: Identifying the convergence settings of juvenile delinquents. *Crime & Delinquency*, *60*(7), 971–998.

Black, D. (1980). *The manners and customs of police*. New York: Academic Press.

Black, D., & Reiss, A. (1970). Police control of juveniles. *American Sociological Review*, *35*, 63-77.

Braga, A. A., Hureau, D. M., & Papachristos, A. V. (2014). Deterring gang-involved gun violence: Measuring the impact of Boston's operation ceasefire on street gang behavior. *Journal of Quantitative Criminology*, 30(1), 113-139.

Braga, A. A., Kennedy, D. M., Waring, E. J., & Piehl, A. M. (2001). Problem-oriented policing, deterrence, and youth violence: An evaluation of Boston's Operation Ceasefire. *Journal of Research in Crime and Delinquency*, 38, 195-225.

Brazil, N. (2020). Effects of public-school closures on crime: The case of the 2013 Chicago mass school closure. *Sociological Science*, *7*(6), 128–151.

Brenner, A. B., Bauermeister, J. A., and Zimmerman, M. A. (2011). Neighborhood variation in adolescent alcohol use: Examination of socio-ecological and social disorganization theories. *Journal of Studies on Alcohol and Drugs.* 72:651–659.

Brantingham, P.L., & Brantingham, P.J. (1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, *13*(1), 3–28.

Brantingham, P., & Brantingham, P. (1995). Criminality of place. *European Journal on Criminal Policy and Research*, *3*(3), 5–26. https://doi.org/10.1007/BF02242925

Brown, B. B. (1990). Peer groups and peer cultures. In S. S. Brown & G. R. Elliott (Eds.), *At the threshold: The developing adolescent* (pp. 171-196). Cambridge, MA: Harvard University Press.

Brunson, R. K., & Miller, J. (2006). Young Black men and urban policing in the United States. *British Journal of Criminology*, 46, 613−640.

Bryant, A. L. and Zimmerman, M. A. (2002). Examining the effects of academic beliefs and behaviors on changes in substance use among urban adolescents. *Journal of Educational Psychology.* 94:621–637.

Burgess, R. L., and Akers, R., L. (1966). A differential association-reinforcement theory of criminal behavior. *Social Problems* 14: 128–46.

Carr, P.J., Napolitano, L., & Keating, J. (2007). We never call the cops and here is why: A qualitative examination of legal cynicism in three Philadelphia neighborhoods. *Criminology*, 45, 445−480.

Carrington, P.J. (2002). Group crime in Canada. *Canadian Journal of Criminology 44:377-15.*

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., & Cho, H. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, *1*(4), 1-4.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). New York, NY, USA: ACM. https://doi.org/10.1145/2939672.2939785

Clarke, R.V. (1980). "Situational" crime prevention: Theory and practice. *British Journal of Criminology*, *20*(2), 136–147.

Clarke, R.V. (1983). Situational crime prevention: Its theoretical basis and practical scope. In M. Tonry and N. Morris (eds.), *Crime and Justice: An Annual Review of Research*. Vol. 4. Chicago, IL: University of Chicago Press.

Clarke, R. V. (1995). Situational crime prevention: Achievements and challenges. In *Building a Safer Society: Strategic Approaches to Crime Prevention*, Michael Tonry & Farrington, David (eds.). Chicago: The University of Chicago Press

Cohen, L. E. and Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review* 44:588-608.

Compas, B. E., Conner-Smith, J. K., Saltzman, H., Thomsen, A. H., & Wadsworth, M. (2001). Coping with stress during childhood and adolescence: Problems, progress, and potential in theory and research. *Psychological Bulletin*, *127*, 87-127.

Conway, K.P., and McCord, J. (2002). A longitudinal examination of the relation between co-offending with violent accomplices and violent crime*. Aggressive Behavior,* 28:97-108.

de Jong, E., Bernasco, W., & Lammers, M. (2019). Situational correlates of adolescent substance use: An improved test of the routine activity theory of deviant behavior. *Journal of Quantitative Criminology*, *36*(4), 823–850.

Dufur, M. J., Parcel, T. L., and McKune, B. A. (2013). Does capital at home matter more than capital at school? The case of adolescent alcohol and marijuana use. *Journal of Drug Issues.* 43(1):85–102.

Durkin, K. F., Wolfe, T. W., and Clark, G. A. (2005). College students and binge drinking: An evaluation of social learning theory. *Sociological Spectrum.* 25:255–272.

Eccles, J., Lord, S., & Buchanan, C. M. (1996). School transitions in early adolescence: What are we doing to our young people? In J. A. Graber, J. Brooks-Gunn, & A. C. Petersen (Eds.), *Transitions through adolescence*. Mahwah, NJ: Lawrence Erlbaum.

Eck, J. E., and Weisburd, D. (1995). Crime places in crime theory. In *Crime Prevention Studies*, eds. John E. Eck and David Weisburd. Monsey, NY: Criminal Justice Press.

Engel, R. S. (2000). The effects of supervisory styles on patrol officer behavior. *Police Quarterly*, 3, 262−293.

Fagan, A. A., Van Horn, M. L., Hawkins, J.D., and Jaki, T. (2013). Differential effects of parental controls on adolescent substance use: For whom is the family most important? *Journal of Quantitative Criminology.* 29(3):347–368.

Farrington, D. (1986). Age and crime. *Crime and Justice (Chicago, Ill.)*, *7*(5), 189–250.

Felson, M., & Gottfredson, M. (1984). Social indicators of adolescent activities near peers and parents. *Journal of Marriage and the Family*, *46*, 709-714.

Felson, M. (2006). *Crime and nature*. Thousand Oaks, CA: SAGE Publications.

Felson, M. (1998). *Crime and everyday life*. Thousand Oaks, CA: Pine Forge Press.

Flacks, S. (2017). The stop and search of minors: A "vital police tool"? *Criminology & Criminal Justice*, *18*(3), 364–384.

Ford, J. A. (2008). Social learning theory and nonmedical prescription drug use among adolescents. *Sociological Spectrum.* 28(3):299–316.

Fratello, J., Rengifo, A., Trone, J., & Velazquez, B. (2013). Coming of age with stop and frisk: Experiences, perceptions, and public safety implications. Retrieved from http://www.vera.org/sites/default/files/resources/downloads/stop-and-frisk_technical-report.pdf

Friedman, W., Lurigio, A. J., Greenleaf, R., & Albertson, S. (2004). Encounters between police officers and youth: The social costs of disrespect. *Journal of Crime and Justice*, *27*(2), 1-26.

Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189–1232.

Geistman, J., & Smith, B. W. (2007). Juvenile attitudes toward the police: A national study. *Journal of Crime and Justice*, 30, 27−52.

Giblin, M. (2002). Using police officers to enhance the supervision of juvenile probationers: An evaluation of the Anchorage CAN program. *Crime and Delinquency*, 48, 116−137.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.

Gottfredson, D. C., Gottfredson, G. D., & Weisman, S. A. (2001). The timing of delinquent behavior and its implications for after-school programs. Criminology & Public Policy, 1(1), 61–86.

Graul, Christian (2016): leafletR: Interactive Web-Maps Based on the Leaflet JavaScript Library. R package version 0.4-0, http://cran.r-project.org/package=leafletR.

Greenberg, D. F. (1977). Delinquency and the age structure of society. *Contemporary Crises*, *1*, 189-223.

Greenberg, D.A. (1979). Delinquency and the age structure of society. In *Criminology Review Yearbook*, eds. Sheldon Messinger and Egon Bittner. Beverly Hills, CA: Sage.

Greenberg, D.A. (1983). Age and crime. In *Encyclopedia of Crime and Justice*, edited by Sanford H. Kadish. New York: Free Press.

Haller, M. H. (1976). Historical roots of police behavior: Chicago 1890-1925. *Law and Society Review*, *10*(2), 303-323.

Helsen, M., Vollebergh, W., & Meeus, W. (2000). Social support from parents and friends and emotional problems in adolescence. *Journal of Youth and Adolescence*, *29*, 319-335.

Hemovich, V., Lac, A., and Crano, W. D. (2011). Understanding early onset drug and alcohol outcomes among youth: The role of family structure, social factors, and interpersonal perceptions of use. *Psychology, Health, and Medicine.* 16(3):249–267.

Hirschi, T. (1969). "Chapter 2: A Control Theory of Delinquency" (pp. 16-26). In *Causes of Delinquency*. Berkeley, CA: University of California Press. ERESERVES

Hoeben, E., & Weerman, F. (2014). Situational conditions and adolescent offending: Does the impact of unstructured socializing depend on its location? *European Journal of Criminology*, *11*(4), 481–499.

Hoeve, M., Geert, J., Stams, J. M., van der Put, C. E., Sermon-Dubas, J., van der Laan, P. H., and R. M. Gerris, J. (2012). A meta-analysis of attachment to parents and delinquency. *Journal of Abnormal Child Psychology.* 40(5):771–785.

Ho, T.K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278–282).

Hurst, Y. G. (2007). Juvenile attitudes toward the police: An examination of rural youth. *Criminal Justice Review*, 32, 121−141.

Hurst, Y. G., & Frank, J. (2000). How kids view cops: The nature of juvenile attitudes toward the police. *Journal of Criminal Justice*, 28, 189−202.

Jackson, A. N. (2013). Assessing the impact of parental drug use, family structure, and environmental conditions on adolescents' self-reported drug use, serious delinquency, and deviant behaviors. *International Journal of Criminology and Sociological Theory.* 6(2):325–354.

Kelly, A. B., O'Flaherty, M., Toumbourou, J. W., Conner, J. P., Hemphill, S. A., and Catalano. R. F. (2011). Gender differences in the impact of families on alcohol use: A lagged longitudinal study of early adolescents. *Addiction.* 106(8):1427–1436.

Kennelly J. (2011). Policing young people as citizens-in-waiting: Legitimacy, spatiality, and governance. *British Journal of Criminology* 51(2): 336–354.

Kochel, T. R., Burruss, G. W., & Weisburd, D. (2015). St. Louis County hot spots in residential areas (SCHIRA) final report: Assessing the effects of hot spots policing strategies on police legitimacy, crime, and collective efficacy. Southern Illinois University.

Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. Journal of cheminformatics, 6(1), 10. https://doi.org/10.1186/1758-2946-6-10

Landis, & Koch, G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lantz, B., & Ruback, R.B. (2017). The relationship between co-offending, age, and experience using a sample of adult burglary offenders. *Journal of Developmental and Life-Course Criminology*, *3*(1), 76–97.

Larson, R. W., Richards, M. H., Moneta, G., Holmbeck, G., & Duckett, E. (1996). Changes in adolescents' daily interactions with their families from ages 10 to 18: Disengagement and transformation. *Developmental Psychology*, *32*, 744-754.

Laub, J. H. and Sampson, R. J. (1993). Turning points in the life course: Why change matters to the study of crime. *Criminology,* 31:301-325.

Lawton, B. A., Taylor, R. B., & Luongo, A. J. (2005). Police officers on drug corners in Philadelphia, drug crime, and violent crime: Intended, diffusion, and displacement impacts. *Justice Quarterly*, 22, 427-451.

Leiber, M., Nalla, M., & Farnworth, M. (1998). Explaining juveniles' attitudes toward the police. *Justice Quarterly*, 15, 151−174.

Liederbach, J. (2007). Controlling Suburban and Small-Town Hoods: An Examination of Police Encounters With Juveniles. *Youth Violence and Juvenile Justice*, *5*(2), 107–124.

Loeber, R., Wei, E., Stouthamer-Loeber, M., Huizinga, D., & Thornberry, T. (1999). Behavioral antecedents to serious and violent juvenile offending: Joint analyses from the Denver Youth Study, Pittsburgh Youth Study, and the Rochester Development Study. *Studies in Crime and Crime Prevention,* 8, 245–263.

Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., Moffitt, T., & Caspi, A. (1998). The development of male offending: Key findings from the first decade of the Pittsburgh Youth Study. *Studies in Crime and Crime Prevention,* 7, 141–172.

Lynch, J. P. (2002). Trends in juvenile violent offending: An analysis of victim survey data. *Juvenile Justice Bulletin*. Washington, DC: U.S. Department of Justice.

MacDonald, J. M., Nicosia, N., & Ukert, B. D. (2018). Do schools cause crime in neighborhoods? Evidence from the opening of schools in Philadelphia. *Journal of Quantitative Criminology*, *34*(3), 717–740.

McCord, J., and Conway, K. (2005). Co-offending and patterns of juvenile crime: Research in brief. *Washington, DC: National Institute of Justice (NCJ 210360)*

Merlo, A.V. & Benekos, P.J. (2010). Is punitive juvenile justice policy declining in the United States? A critique of emergent initiatives. *Youth Justice* 10:3-24.

Meyers, S. M. (2004). *Police encounters with juvenile suspects: Explaining the use of authority and provision of support.* Executive Summary Report submitted to the National Institute of Justice. Available at www.cops.usdoj.gov

Moffitt, T. (1993). Adolescent-limited and life-course persistent anti-social behavior: A developmental taxonomy. *Psychological Review.* 100:674-701. EJournals.

Moffitt, T. E., & Harrington, H. L. (1996). Delinquency: The natural history of antisocial behaviour. In P. A. Silva & W. R. Stanton (Eds.), *From child to adult: The*

Dunedin multidisciplinary health and development study (pp. 163-185). Oxford, UK: Oxford University Press.

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized controlled field trials of predictive policing. *Journal of the American statistical association*, 110(512), 1399-1411.

Moore, & Padavic, I. (2011). Risk Assessment Tools and Racial/Ethnic Disparities in the Juvenile Justice System: Risk Assessment Tools and Racial/Ethnic Disparities. Sociology Compass, 5(10), 850–858.

Mounts, N.S. (2002). Parental management of adolescent peer relationships in context: The role of parenting style. *Journal of Family Psychology* 16:58–69.

Nagin, D. S. (2013). Deterrence in the twenty-first century. *Crime and Justice.* 42(1). 199

Newman, B. M., Lohman, B. J., Newman, P. R., Myers, M. C., & Smith, V. L. (2000). Experiences of urban youth navigating the transition to ninth grade. *Youth & Society*, *31*, 387-416.

Norman, L., & Ford, J. (2015). Adolescent ecstasy use: A test of social bonds and social learning theory. *Deviant Behavior*, *36*(7), 527–538.

Osgood D.W., Wilson J.K., O'Malley P.M., Bachman J.G., & Johnston L.D. (1996). Routine activities and individual deviant behavior. *American Sociological Review,* 61:635–655.

Osgood, D. W., & Anderson. A. L. (2004). Unstructured socializing and rates of delinquency. *Criminology*, *42*, 519-549.

Paoline, E. A., & Terrill, W. (2005). The impact of culture on police traffic stop searches: An analysis of attitudes and behavior. *Policing: An International Journal of Police Strategies and Management*, 28, 455−472.

Perra, O., Fletcher, A., Bonell, C., Higgins, K., and McCrystal, P. (2012). School-related predictors of smoking, drinking, and drug use: Evidence from the Belfast Youth Development Study. *Journal of Adolescence.* 35(2):315–324.

Petersilia, J., Greenwood, P.W., & Lavin, M. (1978). *Criminal Careers of Habitual Felons*. Washington, D.C.: National Institute of Justice.

Piquero, A.R., Farrington, D.P., and Blumstein, A. (2007). *Key Issues in Criminal Careers Research: New Analyses from the Cambridge Study in Delinquent Development.* Cambridge, UK: Cambridge University Press.

Powers, D. M. (2015). What the F-measure doesn't measure: Features, Flaws, Fallacies and Fixes. *arXiv preprint arXiv:1503.06410*.

Ragan, D. T., Osgood, W. D., and Feinberg, M. E. (2014). Friends as a bridge to parental influence: Implications for adolescent alcohol use. *Social Forces.* 92(3):1061–1085.

Ratcliffe, J., Taniguchi, T., Groff, E. R., Wood, J. D. (2011). The Philadelphia foot patrol experiment: A randomized controlled trial of police patrol effectiveness in violent crime hotspots. *Criminology*, 49(3), 795-831.

Rebellon, C. J., Trinkner, R., Van Gundy, K. T., & Cohn, E. S. (2019). No guts, no glory: The influence of risk-taking on adolescent popularity. *Deviant Behavior*, *40*(12), 1464–1479.

Reiss, A., & Farrington, D. (1991). Advancing knowledge about co-offending: Results from a prospective longitudinal survey of London males. *The Journal of Criminal Law & Criminology*, *82*(2), 360–395.

Rengert, G. & Wasilchick, J. (1985). *Suburban Crime: A Time and a Place for Everything.* Springfield, IL: Charles C. Thomas.

Reyes, H. L. M., Foshee, V. A., Bauer, D. J., and Ennett, S. T. (2012). Heavy alcohol use and dating violence perpetration during adolescence: Family, peer, and neighborhood violence as moderators. *Prevention Science.* 13(4):340–349.

Rice, K.J., Smith, W.R. (2002). Socioecological models of automotive theft: Integrating routine activity and social disorganization approaches. *Journal of Research in Crime and Delinquency.* 39(3):304–336

Robins, L. N. (1978). Sturdy childhood predictors of adult antisocial behavior: Replications from longitudinal studies. *Psychological Medicine, 8*, 611–622.

Roman, C., & Lynch, J. (2002). Schools as generators of crime: Routine activities and the sociology of place. ProQuest Dissertations Publishing. Retrieved from http://search.proquest.com/docview/304805774/

Roman C.G. (2005). Routine activities of youth and neighborhood violence: Spatial modeling of place, time, and crime. In: Wang F (ed) *Geographic Information Systems and Crime Analysis. Idea* Group, Hershey, pp 293–310.

Roncek, D.W. (2000). Schools and crime. In: Goldsmith V., McGuire P., Mollenkopf J.H., Ross T.A. (eds) *Analyzing crime patterns: frontiers of practice*. Sage Publications, Thousand Oaks, pp 153–165

Roncek, D.W., Maier, P.A. (1991). Bars, blocks, and crimes revisited: Linking the theory of routine activities to the empiricism of "hot spots". *Criminology*. 29(4):725–753.

Roncek, D. & Lobosco, A. (1983). The effect of high schools on crime in their neighborhoods. *Social Science Quarterly, 64*, 598-613.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sampson, R. J. and Laub, J.H. (1992). Crime and deviance in the life course. *Annual Review of Sociology*, 18:63-84.

Sampson, R. J. and Laub, J. H. (1990). Crime and deviance over the life course: The salience of adult social bonds. *American Sociological Review*, 55:609-627.

Sampson, R. J., Morenoff, J. D., & Earls, F. (1999). Beyond social capital: Spatial dynamics of collective efficacy for children. *American Sociological Review*, *64*(5), 633–660.

Sharp, D. and Atherton, S. (2007). To serve and protect? The experiences of policing in the com- munity of young people from black and other ethnic minority groups. *British Journal of Criminology* 47(5): 746–763.

Shaw, C. R. and McKay, H.D. (1942). *Juvenile delinquency and urban areas; A study of rates of delinquents in relation to differential characteristics of local communities in American cities*. Chicago: University of Chicago Press.

Sickmund, M. and Puzzanchera, C. (eds.). (2014). *Juvenile Offenders and Victims: 2014 National Report*. Pittsburgh, PA: National Center for Juvenile Justice.

Smith, W.R., Frazee, S.G., Davison, E.L. (2000). Furthering the integration of routine activity and social disorganization theories: Small units of analysis and the study of street robbery as a diffusion process. *Criminology*. 38(2):489–524

Steffensmeier, D., & M.D. Harer. (1987). Is the crime rate really falling? *Journal of Research in Crime and Delinquency* 24:23-48.

Steffensmeier, D., & M.D. Harer. (1999). Making sense of recent U.S. crime trends, 1980-96/8: Age-composition effects and other explanations. *Journal of Research in Crime and Delinquency* 36(3).

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. Developmental Review, 28(1), 78–106.

Steinberg, M. P., Ukert, B., & MacDonald, J. M. (2019). Schools as places of crime? Evidence from closing chronically underperforming schools. *Regional Science and Urban Economics*, *77*, 125–140.

Sutherland, E.H. (1947). *Principles of Criminology*, 4th ed. Philadelphia: J. B. Lippincott.

Taylor, T. J., Turner, K. B., Esbensen, F., & Winfree, L. T. (2001). Coppin' an attitude: Attitudinal differences among juveniles toward police. *Journal of Criminal Justice, 29*(4), 295-305.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological), 58*(1), 267-288.

Trucco, E. M., Colder, C. R., and Wieczorek, W. F. (2011). Vulnerability to peer influence: A moderated mediation study of early adolescent alcohol use initiation. *Addictive Behaviors.* 36(7):729–736.

United States Department of Justice, Federal Bureau of Investigation. (2019). *Crime in the United States, 2019.* Retrieved from: https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/tables/table-10/table-10-state-cuts/virginia.xls

U.S. Census Bureau (2019). U.S. Census Bureau Quickfacts: Fairfax County, Virginia. Retrieved from: https://www.census.gov/quickfacts/fairfaxcountyvirginia#qf-headnote-a

Vito, A. G. and Higgins, G.E. (2013). A research note on adolescent steroid use: An examination of social learning theory and self-control theory. *Deviant Behavior.* 34(12):951–960.

Walker, S., & Katz, C. M. (2008). *Police in America: An introduction* (4th ed.). Boston: McGraw-Hill.

Weisburd, D., and Green, L. (1995). Policing drug hot spots: The Jersey City drug market analysis experiment. *Justice Quarterly* 4:711–35.

Weisburd, D., and Mazerolle, L. (2000). Crime and disorder in drug hot spots: Implications for theory and practice in policing. *Police Quarterly* 3:331–49.

Weisburd, D., Morris, N., & Groff, E. (2009). Hot spots of juvenile crime: A longitudinal study of arrest incidents at street segments in Seattle, Washington. *Journal of Quantitative Criminology*, *25*(4), 443–467.

Weisburd, D. (2015). The law of crime concentration and the criminology of place: The law of crime concentration. *Criminology (Beverly Hills)*, *53*(2), 133–157.

Weisburd, D., Bushway, S., Lum, C., & Yang, S. M. (2004). Trajectories of crime at place: A longitudinal study of street segments in the city of Seattle. *Criminology,* 42:283–322.

Weisburd, D., Maher, L., & Sherman, L.W. (1992). Contrasting crime general and crime specific theory: The case of hot spots of crime. In *Advances in Criminological Theory*, eds. Freda Adler and William S. Laufer. New Brunswick, NJ: Trans-action.

Werthman, C., & Piliavin, I. (1967). Gang members and the police. In D. Bordua (Ed.), *The police: Six sociological essays* (pp. 56-98). New York: John Wiley.

West, D.J., and Farrington, D.P. (1977). *The Delinquent Way of Life*. London: Heinemann.

Whitehead, J. T., & Lab, S. P. (1999). *Juvenile justice: An introduction* (3rd ed.). Cincinnati, OH: Anderson.

Wilcox, S. (1973). The Geography of Robbery. In F. Feeney & A. Weir, Eds., *The Prevention and Control of Robbery*, vol.3. Davis, CA: The Center of Administration of Justice, University of California at Davis.

Wilson, J.Q., and Herrnstein, R.J. (1985). *Crime and Human Nature*. New York: Simon & Schuster.

Withrow, B. L., & Bolin, B. (2005). Police protective custody: A systemic predictive model for police decision making and the reduction of referrals. *Policing: An International Journal of Police Strategies and Management*, 28, 473−492.

Wray-Lake, L., Maggs, J.L., Johnston, L. D., Bachman, J. G., O'Malley, P. M., and Schulenberg, J. E. (2012). Associations between community attachments and adolescent substance use in nationally representative samples. *Journal of Adolescent Health.* 51(4):325–331.

BIOGRAPHY

Heather Prince is a Doctoral Candidate in Criminology, Law, and Society at George Mason University. She earned her bachelor's degree in criminology and Sociology from Albright College in 2017, and her Master of Science in criminology from The University of Pennsylvania in 2018. Heather is currently employed at RTI International, as a Research Public Health Analyst. Heather enjoys working with large-scale police and crime data and writing programs in R, as well as mapping and data visualization.