PREDICTING LIFE-COURSE PERSISTENT OFFENDING USING MACHINE LEARNING

____

A Dissertation

Presented to

The Faculty of the Department of Criminal Justice and Criminology

Sam Houston State University

____

In Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

____

by

Gyeongseok Oh

July, 2021

PREDICTING LIFE-COURSE PERSISTENT OFFENDING USING MACHINE LEARNING

by

Gyeongseok Oh

____

APPROVED:

Yan Zhang, PhD
Committee Director

Eric J. Connolly, PhD
Committee Member

Jason Ingram, PhD
Committee Member

Phillip Lyons, PhD
Dean, College of Criminal Justice

DEDICATION

I dedicate this dissertation to my parents and my daughters, Aileen and Hannah.

ABSTRACT

Oh, Gyeongseok, *Predicting life-course-persistent offending using machine learning*. Doctor of Philosophy (Criminal Justice), July 2021, Sam Houston State University, Huntsville, Texas.

The current study investigated the predictive ability of Life-Course-Persistent (LCP) offenders using Machine Learning techniques. Drawing on the National Longitudinal Survey of Youth 1997, LCP and adolescent limited offenders are identified by the latent class growth analysis. Using seven types of Machine Learning techniques, the LCP offenders are predicted by risk factors verified by previous empirical studies. The results of predictive modeling reveal that the Machine Learning-based prediction of LCP offenders significantly outperforms the conventional parametric statistical analysis, logistic regression. Most of all, the predictive ability of Random Forests and Deep Learning model show a more effective forecasting ability than other Machine Learning-based modeling and logistic regression analysis.

KEY WORDS: Machine learning, Life-course-persistent offender, Predictive modeling, Developmental criminology

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Page

CHAPTER I

Introduction

Statement of Problem

The Science Council of the United Kingdom defines science as “the pursuit and application of knowledge and *understanding* of the natural and social world following a systematic methodology based on evidence.” As described in the definition of science published by the Science Council, many social scientists have highly focused on the *understanding* function of science, thus, the pursuit of social science seemed limited to explain the social world (see Weber, 1978). Similarly, in Weber (1978)’s definition of sociology, the principal role of social science was described as understanding and explaining:

> *Sociology is a science concerning itself with the interpretive understanding of social action and thereby with a causal explanation of its course and consequences.* (Weber, 1978, p. 4)

Accordingly, to examine and verify the causal explanations of social phenomena, social science relied on the null hypothesis significance testing process in which theory-based hypotheses are statistically examined using observed data (see Bushway, Sweeten, & Wilson, 2006). Therefore, it is not surprising that contemporary social scientific research majorly focuses on testing theories and explaining the causal mechanisms between cause and effect. However, it should be pointed out that science is not exclusively seeking out the *explanation* of phenomena in the world based on the observation of the past. There is another pillar of science, the *prediction* of the future.

Philosophers of science keep arguing that the central purposes of science are twofold, *prediction,* and *explanation* (see Purtill, 1970). Dubin (1978) wrote in his book:

> *Theories of social and human behavior address themselves to two distinct goals of science: (1) prediction and (2) understanding. (*Dubin, 1978, p. 9)

In an ideal world, both goals of science should be equally valued and dealt with by researchers. However, contemporary customs of social science, including criminology and criminal justice, nearly neglected the importance of predicting social and individuals acts (see Hofman, Sharma, & Watts, 2017). Historically, psychology focuses on explaining human behavior based on existing theories (Yarkoni & Westfall, 2017), and social science generally put less importance on prediction relative to explanation due to the inherent complexity of social systems (Hofman, Sharma, & Watts, 2017). Although some clinical assessment tools were used to predict behaviors, most psychological and social scientists have put more resources on explaining human behaviors rather than predicting the behaviors (Yarkoni & Westfall, 2017). Criminology and the criminal justice field also was not the exception to this traditional approach of social science. A plethora of criminology theories have emerged and statistical models were applied to test hypotheses on relationships between variables by utilizing null hypothesis significance testing process, very few studies attempted to predict the criminality of individuals or crime trends in the future (see Bushway, Sweeten, & Wilson, 2006).

Why contemporary scientific research in criminal justice heavily focused on understanding causal mechanisms other than predicting the outcome? Statisticians and computer scientists responded to the question that the conflation between explanation and prediction (see Shmueli, 2010), and the lack of knowledge regarding predictive modeling (see Hofman, Sharma, & Watts, 2017) hindered scientists from emphasizing the role of prediction. While prediction and explanation are two discrete concepts (see Scriven, 1959), there was a long history of the conflation of prediction and explanation by assuming two concepts equivalent in the field of science (Hempel & Oppenheim, 1948). In particular, once the causal mechanisms were found using *explanatory modeling* based on observed data, scientists believed that predicting the outcome can be achieved by reversing the process of causal inferences. Briefly speaking, it was believed that explanation and prediction were two sides of a coin. Accordingly, scientists have assumed that the high explanatory power of their statistical modeling means high predictive power in general (Shumueli, 2010).

Another reason for deemphasizing the importance of predicting social and individual acts is derived from the lack of knowledge regarding recently developed *predictive modeling* techniques (see Hofman et al., 2017). The conventional statistical modeling was mostly designed to find whether a particular effect in an idealized model is statistically significant (Hofman, et al., 2017). On the contrary, the scientific research methods for prediction require distinct steps of the statistical modeling process which is different from the conventional approach of scientific research. Shmueli (2011) defined this application of conventional statistical models to data for testing causal hypotheses as *explanatory modeling*, whereas *predictive modeling* refers to the process of predicting future observations by applying the statistical model or data mining algorithms (Shmueli, 2010).

Conventional scientific methods rooted in *explanatory modeling* have provided scholars and practitioners in the field of criminology and criminal justice with plenty of explanations on the causes of crime and antisocial behaviors. During the past few decades, scholars of criminology and the criminal justice field have developed numerous theories to explain individual antisocial/criminal behavior and to find the etiologies of those behaviors. Based on the evidence of causal inferences drawn from explanatory studies, administrators and policymakers developed a variety of crime prevention and treatment programs to deal with the crime problems of society. Some programs have targeted removing individual-level causes of delinquency or crime by providing juvenile offenders with early treatment programs (see Piquero, Farrington, Welsh, Tremblay, Jennings, 2008) or awareness programs (see Petrosino, Turpin-Petrosino, & Buehler, 2013). Others have focused on social contextual causes of crime to reduce crime rates, such as crime prevention through environmental design (Jeffery, 1971), neighborhood crime watch (Bennett, Farrington, & Holloway, 2008), or problem-oriented policing (see Hinkle, Weisburd, Telep, & Peterson, 2020). The results of these programs however are mixed. For instance, the family and parent training programs were found to be effective in reducing the chances of behavioral problems and recidivism (Piquero et al., 2008), while the juvenile awareness programs rather increased delinquency of individuals overall (Petrosino et al., 2013). In addition, in the systematic review of crime prevention programs of Poyner (1993), it is found that prevention programs via social and community services were mostly ineffective, while community-oriented policing and environmental designing programs showed significant decreases in crime rates of the targeted areas.

As Gottfredson (1987) pointed out, prediction is a required step if one seeks to control criminal behavior. However, the process of validating the *prediction* has been overlooked by researchers and policymakers. I suspect that weak impacts and the failure of some treatment programs in the criminology and criminal justice field may be influenced by this missing link. I argue that the missing link, predicting their criminal behavior using theory-based factors, should be accomplished to undergird the foundation of the policy and treatment programs.

A similar argument already has been made by Gottfredson a few decades ago. In the late 1980s, scholars in the field of criminal justice, have highlighted the importance of prediction in behavioral science, criminological research, and criminal justice decision making (see Gottfredson, 1987). Gottfredson (1987) argued that the prediction of criminal involvement of individuals in the future can be achieved by a predictive approach, thus criminal justice agencies may focus on reducing crime rates. In addition, estimating the probability of recidivism of offenders may assist judges and probational officers when making the judicial decision (Gottfredson, 1987). As consequence, numerous types of structured risk assessment tools designed to predict the probability of violence or criminal offending have been developed. According to a recent systematic review (Fazel et al., 2012), there were more than 150 structured risk assessment tools including Level of Service Inventory-Revised (LSI-R), Psychopathy Checklist-Revised (PCL-R), Historical, Clinical, Risk management-20 (HCR-20), and Structured Assessment of Violence Risk in Youth (SAVRY) in the world. Although the risk assessment tools are used when making decisions of parole in more than 20 states in the United States, less than two hundred researches were found of which the predictive ability of the risk assessment tools was validated. Moreover, the predictive researches majorly have focused on the *explanatory* approach using conventional parametric statistical models such as logistic regression, while few recent studies started to adopt nonparametric and algorithmic predictive modeling (Fazel et al., 2012). All in all, it is too early to state that scholars in criminology and criminal justice have built the fundamentals to utilize *predictive modeling* in forecasting individual criminality in the future.

In addition to the prediction of individual problems of behavior, policing scholars and administrators recently have begun to have interests in the predictive policing strategy. With the development of quantitative analysis techniques and the increase in the amount of available data, it became available to forecast when and where crime will occur (Uchida, 2009). Meijer and his colleagues (2019) defined predictive policing as *the collection and analysis of data about previous crimes for identification and statistical prediction of individuals or geospatial areas with an increased probability of criminal activity to help developing policing intervention and prevention strategies and tactics* (Meijer et al., 2019, p. 1033). One of the most outstanding predictive policing techniques is Risk Terrain Modeling (RTM). Beyond the retrospective analysis of hot spots of crime, RTM is designed to forecast certain types of crime using place-based risk factors (Caplan, Kennedy, & Miller, 2010). Rooted on the opportunity theory of crime (Cohen, Kluegel, & Land, 1981), RTM identifies all place-based risk factors of crime in a Geographic Information System (GIS) and provides a risk terrain map by combining the results of risk evaluation. Since the purpose of RTM is not to analyze the hot spots of crime incidents ex-ante, but to provide the crime forecasting map ex-post, a *predictive modeling* approach is appropriate to evaluate the effectiveness of the forecasting model. Hence, Caplan and his colleagues (2010) examined the predictive accuracy of RTM for shooting and found that the overall accuracy of forecasting high-risk places for the shooting is significantly higher than that of conventional hot-spot analysis.

Several police departments in big cities such as New York and Los Angeles also adopted the predictive policing software in practice and that results in crime reduction (Levine, Tisch, Tasso, & Joy, 2017; Mohler et al., 2015). For instance, since 2012, New York Police Department has adopted the Domain Awareness System (DAS) to forecast certain types of criminal behavior including robbery, shooting, and burglary using the previous records of crime incidents, 911 calls, CCTVs, environmental sensors, and license plate readers. A recent study evaluated the effectiveness of DAS and revealed that the accuracy of predicting crime incidents is significantly higher than hot-spot analysis and the overall crime index of the city of New York decreased by 6 percent after adopting DAS (Levine, Tisch, Tasso, & Joy, 2017). However, the empirical studies mostly focused on evaluating the crime reduction effects of the predictive policing techniques instead of validating the performance of prediction (see Mohler et al., 2015; Hunt, Saunders, & Hollywood, 2014).

Notwithstanding the importance of predictive roles of science, *predictive modeling* should be more emphasized in the era of Big Data since the conventional explanatory methodologies would be jeopardized with the emergence of massive data. With the exponential growth of internet connectivity and the proliferation of personal IT devices across the world, the quantity of data generated by human activity has exploded during the past two decades. Alongside the growth of the amount of data, industrial parties highlighted the potential application of big data to analyzing patterns of human behaviors and predicting individuals' decisions. On the contrary, social scientists generally cast doubts on the use of big data for research purposes because the conventional *explanatory modeling* which adopts the hypothesis testing process may yield the wrong conclusions when using big data. To be specific, with large size of samples, the probability of Type 1 error – false discoveries – will automatically increase, in other words, almost all research hypotheses can be falsely accepted if researchers only rely on conventional explanatory approaches (see Lohr, 2012). Indeed, some scholars alarm that Big Data would bring about the end of theory due to excessive reliance on the inductive reasoning approach (Anderson, 2008). However, the concerns of social scientists are largely caused by the reliance on *explanatory modeling* and by the ignorance of the predictive role of science. At this point, I agree with the argument that changes in the epistemological approach using *predictive modeling* should be more emphasized in the era of Big Data (see Kitchin, 2014). It is time to amalgamating the conventional deductive methodologies and inductive approaches using *predictive modeling* (see Kitchin, 2014). If scientists begin to regard two approaches, *predictive and explanatory modeling*, as equally valued, the new possibility of research and theory that were unavailable due to the paucity of data will be rewarded (see Shah, Cappella, Neuman, 2015).

To fill this void of *predictive modeling* in the field of criminology and criminal justice, I argue that Machine Learning techniques should be emphasized. Machine Learning (hereafter ML) is the statistical model that allows computers to perform a specific task by learning patterns and inferences using training data without being explicitly programmed (Berk, 2008; Samuel, 1959). In *explanatory modeling*, the parametric statistical model is commonly used because the terms of statistical model functions are easily linked to the underlying theoretical model, thus the statistical model is interpretable using the names of each variable (Shmueli, 2010). For instance, one can interpret the following equation of ordinary linear regression analysis with “One unit increases of the independent variable is significantly associated with the β unit changes of dependent variable”. On the other hand, algorithmic methods and nonparametric models are not generally adopted due to the difficulty in interpreting the model in the *explanatory modeling*.

$$y_i = \alpha + \beta x_i + \varepsilon$$

On the contrary, the top priority of *predictive modeling* is not the interpretability but the accurate prediction of the observations using complex and sometimes unknown functions. Accordingly, both algorithmic approach and statistical models can be widely used in *predictive modeling* along with conventional parametric models regardless of their interpretability (Breiman, 2001). The algorithms of ML are designed to build the most efficient and accurate *predictive mathematical model* via mining the training datasets (see Shmueli, 2011). Since the idea of machine learning was suggested by Samuel (1959), the statistical learning algorithm has been improved along with the development of computational statistics.

Contemporary machine learning techniques even make computers create artificial neural networks that imitate the networks of neurons in the human brain to find latent patterns and inferences without the instructions and supervision of the human being. From the first machine learning program playing checker games (Samuel, 1959) to the latest deep learning program playing GO games (see Silver et al., 2017), scientists witnessed how the ability of artificial intelligence in learning patterns and solving problems has improved dramatically.

Recently, combined with the extensive growth of the size of available data, contemporary ML techniques keep developing their algorithms to improve the predictive ability. As the capability of big data and machine learning techniques progressed in exploring hidden patterns in data and predicting observations, scholars in many fields of science also paid vigorous attention to building ML-based *predictive modeling*. Machine learning was applied to detect and diagnose disease (Sajda, 2006), to generate and test theories of molecular and material science (Butler et al., 2018), to predict the stock market (Patel et al., 2015), or to analyze and annotate a wide variety of genomic sequence elements (Libbrecht & Noble, 2015).

Some scholars across social science begin to advocate hybrid methods that integrate computational algorithms such as machine learning with conventional research methods (Burscher et al., 2015; Park et al., 2015; Zamith & Lewis, 2015). Shah et al. (2015) called this new hybrid method *computational social science* and defined it as:

> *a subcategory of work on big data (1) that uses large, complex datasets; (2) that frequently involves with "naturally occurring" social and digital media sources and other electronic databases; (3) that uses computational or algorithmic solutions to generate patterns and inferences from these data; and (4) that is applicable to social theory in a variety of domains from the study of mass opinion to public health, from examinations of political events to social movements.* (Shah et al., 2015, p.7)

Correspondingly, in the criminology and criminal justice field, few attempts have been made to build *predictive models* forecasting individual involvement of violent or criminal behaviors using risk assessment tools (Gardner et al., 1996; Derogatis & Melisaratos, 1983; Steadman et al., 2000; Silver et al., 2000; Berk, Kriegler, & Beak, 2005; Rosenfeld & Lewis, 2005; Berk, Sherman, Barnes, & Ahlman, 2009; Neuilly et al., 2011; Berk & Bleich, 2014; Pflueger et al., 2015; Holleran & Stout, 2017). Especially in the 1990s, ML techniques received attention from psychiatrists to improve practitioners' screening ability of risky mentally disordered individuals (see Gardner et al., 1996; Steadman et al., 2000, Pflueger et al., 2015). To clinical practitioners, the forecasting results of ML are more incisive and supportive to identify high-risk patients comparing to the conventional risk assessment processes, such as summing up the scores of items of the risk assessment tool. In similar ways, several attempts have been also made to apply ML to forecasting recidivism and rule-breaking behaviors of offenders (Berk, Kriegler, and Beak, 2005; Rosenfeld & Lewis, 2005; Berk, Sherman, Barnes, & Ahlman, 2009; Neuilly et al., 2011; Berk & Bleich, 2014). However, it is still difficult to ascertain whether predictive models using ML outperform the model using conventional statistical methods according to the results of empirical studies (Duwe & Kim, 2018; Neuilly et al., 2011; Ozkan et al., 2020). On the other hand, *predictive models* using ML have been found to be more accurate/effective than traditional regression models (Rosenfeld & Lewis, 2005; Berk, Sherman, Barnes, & Ahlman, 2009; Neuilly et al., 2011; Holleran & Stout, 2017).

Regardless of the performance of the predictive models of forecasting recidivism or inmate's misconduct using ML, however, no attempts have been made to build predictive models of forecasting who consistently commit criminal offending and who desist from those behaviors using ML. Indeed, studies mentioned above applied ML techniques to building *predictive models* of recidivism who already incarcerated (Berk, Kriegler, and Beak, 2005; Rosenfeld & Lewis, 2005; Berk, Sherman, Barnes, & Ahlman, 2009), or violent behaviors of mental disorder patients (Gardner et al., 1996; Steadman et al., 2000, Pflueger et al., 2015) rather than predicting the chronic offenders or offenders who desist from crime. According to the developmental perspective of crime, certain risk or preventive factors make individuals escaping from criminal involvement or remaining persistent criminal offenders (see Moffitt, 1993; Sampson & Laub, 2005). While a plethora of empirical studies revealed that several individual or environmental-level variables are statistically significantly associated with the LCP and AL offenders (Bergman & Andershed, 2009; Moffitt, Caspi, Rutter, & Silva, 2001; Dubow et al., 2014), the accuracy of a predictive model using the risk/preventive factors have yet been examined.

In this regard, the current study first attempts to develop the *predictive model* using ML techniques to identify life-course-persistent offenders who would be more likely to engage in criminal behavior in adulthoods. From the perspective of the risk factor prevention paradigm, the early identification of risky juveniles and proper treatments to those of high risk is the foremost task to prevent persistent criminal involvement of individuals (see Farrington, 2000). The evidence of several longitudinal studies of criminal involvement of individuals undergirds the argument by revealing that only a few numbers of juveniles remain behaving antisocially while a majority of delinquent juveniles stops engaging in criminal behaviors by the early twenties (see

Jolliffe et al., 2017). Despite abundant evidence on the causal relationships between risk /protective factors on criminal involvement according to existing criminological theories, the overarching *predictive model* to identify individuals who would engage in criminal behaviors in the future based on the individual and environmental characteristics of their childhoods has yet been established. As shown in the most recent systematic review of the life-course theory of crime (Jolliffe et al., 2012), all studies applied the *explanatory modeling* designed to find the significant factors that increase the odds of the likelihood of being life-course-persistent offending. However, if the goal of prospective longitudinal studies is to examine the ability of the risk assessment tools of predicting life-course-persistent offenders, the research is sought to examine the criteria based on *predictive modeling*. Stated differently, instead of finding significant coefficients of regression analysis, the predictive performance parameters should be examined. Therefore, the first goal of the current study is to examine the predictive ability of life-course-persistent offenders using risk factors according to the *predictive modeling* approach.

Another goal of the current study is to apply nonparametric statistical models and algorithmic methods to build *predictive models* of LCP offenders. With the combination of the ML-based *predictive model*, it is expected that the predictive ability of the ML-based model of criminal offenses will be higher than that of conventional statistical methods. Hence, the current study focuses on building a predictive model using both ML-based techniques and conventional statistical methods that forecast lifetime persistent criminals using risk/protective factors at the earlier stage of life. Building upon existing criminological theories, individual or environmental risk/protective factors will be included in the *predictive model*. Afterward, several predictive models such as tree-based and neural networks system-based approaches will be developed. It is expected that the ML-based classification tool will be an attractive method if the advanced forecasting model of future criminal involvement using theory-based risk/preventive factors is proven to be effective.

Purpose of the Study

The causes of individual criminality and the differences in life trajectories of problematic behaviors have been uncovered by explanatory modeling for the past decades. Several theories have built the fundamentals to explain criminal offending and the life-course-persistent offenders and observed datasets proved the validity of theories. However, the predictive ability of the developmental theory has yet been examined using predictive modeling approaches. Hence, it is urged to conduct studies investigating the predictive performance of the developmental theory of crime. In doing so, most advanced predictive modeling techniques based on algorithmic and non-parametric models that were unable to be applied in the explanatory modeling should also be employed.

To this end, the purpose of the present study is twofold. First, it examines the validity of the predicting model of life-course-persistent offenders using theory-based risk factors according to the developmental theory of crime. Secondly, it utilizes advanced Machine Learning techniques along with conventional parametric statistical methods to build the predicting models and compare their predictive ability between them. The results of the study are expected to validate the developmental theory of crime and provide substantive supports for risk factor assessment tools using Machine Learning techniques.

Research Questions

Research Question #1: Can the risk and preventive factors of criminal offending based on several theoretical backgrounds predict the life-course-persistent offenders?

Research Question #2: Can machine learning techniques yield better performance in predicting life-course-persistent offenders than conventional statistical methods?

The following chapter provides an overview of the literature on the developmental theory of crime and the application of machine learning methods in predicting the criminality of individuals. Chapter three outlines the source of data, operationalization of variables, and analytic strategy of the study. Chapter four provides results. Chapter five presents a discussion of findings and policy implications, limitations, and guidance to the future study.

CHAPTER II

Literature Review

"Man by nature desires to understand the universe in which he lives"

Aristotle, Unknown

Developmental Perspectives and Risk Factor Prevention Paradigm

Delinquent behaviors of juveniles sometimes are considered as early signs of criminality, in the meantime, delinquency is a quite common behavioral characteristic of adolescence. The developmental perspective of crime contends that delinquency is rather a method to cope with the stress derived from the gap between the current social status and the expectation of juveniles. Oftentimes, it is witnessed that most delinquent adolescents desist antisocial behaviors by the early twenties. According to surveys of a nationally representative sample in the US, nearly 70 percent of juveniles have experienced delinquent behaviors, while only 24% of them have had an experience of arrest or contact with police (Huizinga & Elliot, 1987). For that reason, the developmental perspective of crime highlighted the prevalence of criminal behaviors during adolescence and focused on the discrete characteristics between two groups, who desist crime and who failed to desist criminal involvement at the end of their puberty (see Moffitt, 1993). Therefore, to prevent crimes, the risk factor prevention paradigm emphasizes the early identification of high-risk juveniles who are more likely to become life-course-persistent criminals for the proper treatment. To this end, building forecasting models to predict persistent criminal involvement using risk factors at the early stage of life is the most required job under the risk factor prevention paradigm.

To build robust and plausible predictive tools of life-course-persistent offenders, it must be founded on scientific theories and evidence. In a similar vein, the identification of risk/preventive factors of criminal involvement should be founded on criminological theories and empirical evidence of those theories. Among a plethora of criminological theories that explain different types of crime such as, violent or white-collar crimes, some theories are applicable to general criminal involvement. Thornberry and his colleagues (2012) categorized the general theory of crime into five dimensions; static, dynamic, social-psychological, developmental psychopathology, and biopsychosocial perspective. Each of the five dimensions of the theoretical perspective of criminal involvement will be discussed followed by results of empirical studies of each theory.

Static Theories

The first general theory of crime that explains consistent criminal or antisocial behaviors is Gottfredson and Hirschi's (1990) self-control theory. Self-control theory, also known as the "General Theory of crime", argues that individuals who have a low ability of self-control are more likely to commit crime because they are prone to be risk-taking, impulsive, and insensitive to others. Self-control theory also highlighted the intra-stability of low self-control across an individual's lifetime. In their work (Gottfredson & Hirschi, 1990), they insisted that parenting effectiveness is the key to grow children's ability of self-control and the ability of self-control is determined at the very early stage of life (Age of 8). Several empirical studies have examined the association between low self-control and antisocial behaviors and delivered supportive findings (see Pratt & Cullen, 2000). Furthermore, longitudinal studies have revealed that self-control does not significantly change over time (Beaver & Wright, 2007; Hay & Forrest, 2006), thus the effect of low self-control traits may continue over the life course. For instance, Beaver and Wright (2007) examined the stability of self-control from kindergarten to the first grade of elementary school using the Early Childhood Longitudinal Study which comprised 17,000 kindergarten students. Their findings also revealed that low self-control is highly stable in early childhood.

Moffitt (1993) named individuals who engage in consistent criminal or antisocial behaviors as life-course-persistent criminals. According to Moffitt's (1993) argument, individuals who hold low cognitive ability, poor self-control, irritability, and a high level of activity are expected to engage in criminal offending across the life course. Moreover, Moffitt (1993) argued that individuals lacking in controlling their impulse and sympathy at a younger age are more likely to be drawn to criminogenic environments such as delinquent peer groups, or school failures. Consequently, the cumulative continuity of a hazardous lifestyle leads to criminal involvement at later stages of life. As an example, Wright, Caspi, Moffitt, and Silva (1999) examined the cumulative effects of low self-control in childhood and found that the children who present low self-control are more likely to engage in criminal behavior, moreover, the association was mediated by weak social bonds.

Dynamic Theories

The developmental perspective of crime also seeks to explain the relationship between criminal behaviors and the social environments across their life course. As mentioned above, some factors of static theories may lead individuals to certain environments and produce cumulative effects. However, other dynamics of environments, such as parenting, family, school, friends, and job also independently divert the lifestyles of individuals to the chronic offenders. Dynamic theories suppose that human behavior is not an outcome of static characteristics, rather the process of interacting with social environments. Social learning (Akers & Jennings, 1998), differential association (Sutherland, 1956), and life-course theory (Sampson & Laub, 2003) adopt this sociogenetic perspective. Differential association theory (Sutherland, 1956) argues that delinquency is an expression of an individual's delinquent values that are learned by interacting with intimate others. Through the repetitive process of behaving delinquency and justification for the behavior, belief, and value of the social norm are established, as a result, the individuals with a low standard of the anti-social norm will engage in criminal behavior. Social learning theory (Burgess & Akers, 1966; Akers & Jennings, 1998) adopts the proposition of Differential Association theory and expanded the theory by embracing instrumental conditioning. Instrumental conditioning also called operant conditioning, is a type of learning process that occurs through positive and negative reinforcement responding to certain behaviors (Skinner, 2019). Applying the theoretical framework of operant conditioning, Akers (1985) argues that crime is also an outcome of established value that is learned through the process of positive rewards and the absence of negative stimuli for delinquent behaviors. Social learning theory is supported by numerous empirical evidence explaining several types of crime such as computer crimes (Skinner & Fream, 1997), terrorism (Akers & Silverman, 2004), juvenile delinquency (Warr & Stafford, 1991), and criminal conviction (Rowe & Farrington, 1997). For instance, in the study of Skinner and Fream (1997), individuals learned illegal computer activities through their intimate groups such as peers, family, and teachers. Moreover, their computer crimes are reinforced by experiencing a lack of penalties.

Along with the learning processes of social norms and beliefs towards crime, the life-course theory of crime focuses on other social and individual dynamics that may cumulatively impact individual trajectories of offending. The life-course theory of crime argues that several social environments and individual traits may lead the individual to start antisocial or criminal behaviors, to continue their behaviors, or to desist from the behaviors (Sampson & Laub, 1993). The theory assumes that ineffective parenting, poverty, school failure, peer delinquency, and negative individual trait are associated with the early onset of delinquency (Thornberry & Krohn, 2005). As a consequence of the delinquent behavior of initial offending, individuals may be alienated from positive social environments and become close to negative environments such as school failure, delinquent peers, and parental rejection (Thornberry, 1987). While the risky environments cumulatively push the individuals to the crime-prone lifestyle, some positive environments have individuals desist from engaging in crime (Sampson & Laub, 1993, Thornberry, 1987). Offenders whose bonds to conventional society are re-established through having children, marriage, or stable job experience are more likely to have strong social bonds and are less likely to have the opportunity for criminal behavior. As a consequence of these positive changes in environments, individuals may escape from the downward spiral of criminal careers. The findings of some longitudinal studies supported the arguments of life-course theory. One of the prominent longitudinal surveys, the Cambridge Study in Delinquent Development (CSDD) was utilized several times for investigating the impact of risky environments including family, socioeconomic, other psychological dynamic factors on the early onset of offending (Zara & Farrington, 2013). A series of studies examined the effects of static and dynamic risk factors on delinquency and criminal involvement across their lifetime using CSDD data. Farrington (2001) summarized the findings drawn from the studies using SDD and concluded that the six most important predictors of offense in the future were found at age 8 - 10; Antisocial child behavior, hyperactivity-impulsivity-attention, low intelligence-low school achievement, family criminality, family poverty, and poor parenting.

Social Psychological Theories

While static and dynamic theories consider behavioral problems as deterministic outcomes of deterministic outcomes caused by risk factors, social psychological theories focus on the concept of human agency. According to the idea of human agency, the levels of ability to interact with surrounding environments vary individually (Matsueda & Heimer, 1997). The social psychological theory posits that human behaviors are the results of their choices that rely on their capacity of adjusting to their environment. For instance, social-psychological theories argue that cognitive bias distorts the intention of others, therefore, individuals with cognitive distortion tend to respond more aggressively than others in the same stimulation from the environments (Dodge, 2006). While their aggressive tendency is relatively consistent across their lifetime, it is also argued that hooks for changes may transform their cognitive bias (Giordano et al., 2002). As an example, positive social networks such as prosocial spouses and social experiences are sources of altering their cognitive distortion.

Developmental Psychopathology and Biopsychosocial Perspective

Neither static traits of individuals nor social environments play independent roles in the developmental process of individuals. During the development of individuals from their early childhood to adulthood, genes, experiences, and the developments of the brain interact with static traits and dynamic environments. Further, these complex interactions are related to behavioral problems.

Developmental psychopathology undertakes the genetic variation between individuals and their interplay with changes in environments. For instance, certain genetic polymorphisms may push individuals to negative environments or accelerate the impact of negative environments (Rutter, Moffitt, & Caspi, 2006). Along with genetic variation, the biopsychosocial perspective also focuses on the structure and function of the brain and its developments across its lifetime. For instance, it is found that the structure and function of the brain significantly differ between offenders and nonoffenders (Raine, 1993) and criminal psychopaths and normal individuals (Blair, Mitchell, & Blair, 2005).

Theory-Based Risk Factors for Criminal Behavior

There have been accumulative efforts of researchers to identify and categorize the risk and protective factors based on the theoretical framework explaining criminal behaviors. The main purposes of these risk assessment tools are to identify the high-risk individuals and to prevent future crimes through proper interventions. Therefore, from the perspective of practitioners, it is expected that criminology and criminal justice studies provide them with successful tools for predicting the criminal involvement of individuals in the future.

Since the actuarial assessment was suggested by Burgess (1928), several sophisticated risk assessment scales have been developed based on empirical evidence. As an example, one of the outstanding risk assessment scales is Andrews, Bonta, and Wormith's (2000) Level of Service / Case Management Inventory (LS/CMI). LS/CMI consists of 8 general risk/need factors and 7 specific risk factors. The general risk/need factors consist of eight categories; criminal history, education/employment, family/marital, leisure/recreation, companions, alcohol/drug problem, pro-criminal attitude/orientation, and antisocial pattern. The other class of LS/CMI consists of seven specific risk factors related to personal problems with criminogenic potential and the history of perpetration such as psychopathy, anger management deficits, poor social skills, underachievement, sexual assault, physical assault, and gang participation. Following the preexisting framework of risk assessment tools of criminal behavior and delinquency, the current study categorizes the risk/protective factors into three dimensions: Familial characteristics, individual characteristics, and peer/environmental characteristics.

Individual Factors

A variety of biological and psychological factors are found to be related to antisocial behaviors and youth violence (Wright, Caspi, Moffitt, & Silva, 1999). These include individual characteristics that may increase the probability of being exposed to risky environments or may lead to failure of controlling their impulse of giving out emotional drives. As discussed above, an individual's trait or emotional problems are found to be strongly related to later violence or criminal behaviors. Low self-control (Wright, Caspi, Moffitt, & Silva, 1999), hyperactivity (Jolliffe & Farrington, 2009), low behavioral inhibition, and aggressiveness are included in this category. Low self-control is one of the most frequently examined risk factors of antisocial behavior. Wright and his colleagues (1999) examined the long-term effect of self-control in childhood on criminal behavior at the later stage of life using a longitudinal study in Dunedin, New Zealand. Their study revealed that low self-control in childhood increased the likelihood of engaging in criminal behavior in adulthood, while the association was mediated by social bonds. Although Gottfredson and Hirschi (1990) viewed the trait of hyperactivity and impulsivity as a manifestation of low self-control, several empirical studies have also found the independent influence of impulsiveness and hyperactivity disorder on later violence and crime (Jolliffe & Farrington, 2009; Pratt, Cullen, Blevins, Daigle, & Unnever, 2002). Jolliffe and Farrington (2009) analyzed six longitudinal studies and found that the impulsiveness of the individual in childhood significantly predicted violence in the future. Similarly, Pratt and his colleagues' (2002) meta-analysis on Attention Deficit Hyperactivity Disorder (ADHD) and crime also found that the strong influence of ADHD on criminal and delinquent behavior. Further, it is found that the influence was constant across twenty different empirical studies which are adopted to their meta-analysis.

General Strain Theory contends that individuals engage in crime and antisocial behaviors to alleviate their negative emotionality emanated from strain-inducing experiences (Agnew, 1992). Lots of empirical studies support that the directive or indirective strain-inducing experience increases the likelihood of crime and deviant behaviors, and the association was mediated by negative emotionality (Agnew & White, 1992; Broidy, 2001; Jang & Johnson, 2003; Rebellon, Manasse, Van Gundy, & Cohn, 2012, Oh & Connolly, 2019, Aseltine, Gore, & Gordon, 2000). For instance, longitudinal studies revealed that negative life events such as life stress, family conflict, peer conflict (Aseltine et al., 2000), and bullying victimization (Oh & Connolly, 2019) increased the negative emotionality which in turn increased the likelihood of engaging in crime or deviant behaviors. As wells as strain-inducing life experiences, Agnew and his colleagues also focused on the personality traits that make individuals are conducive to become aggressive, hostile, and angry (Agnew, Brezina, Wright, & Cullen, 2002). Findings of their study revealed that the negative emotionality and low constraint conditions the effect of strain on delinquent behavior, such that who has negative personality trait are more likely to react to strain-inducing life experiences with delinquency.

Family Factors

Parenting and familial characteristics are the most important environments during the child development period. The influence of family and parenting styles on children's delinquency and crime in their adulthood is threefold. The first domain of the family and parents' influence is related to the social control aspect. As argued by Hirschi (1969), individuals with high attachment to their parents are less likely to engage in antisocial behaviors because they tend to avoid losing emotional bonds to their parents. Abundant empirical evidence supports the link between parental attachment and antisocial behaviors (Wiatrowski, Griswold, & Roberts, 1981; Cernkovich & Giordano, 1987; Loeber & Loeber, 1986). In addition to the lack of social bonds, Gottfredson and Hirschi (1990) also highlighted the negative influence of parenting habits and social bonds on the level of self-control of their children. The general theory of crime (Gottfredson &

Hirschi, 1990) insists that the lack of parental supervision and parental rejection bring about low self-control of their children, which in turn increases the likelihood of showing delinquency and antisocial behaviors. Empirical evidence also supports the effect of poor parental supervision on delinquency (Jang & Smith, 1997), however, the mediating role of self-control on the association between parenting and delinquency/crime is still controversial (see Burt, Simons, & Simons, 2006). The status of empirical evidence on the impact of parental attachment on the problem behaviors of their children is robust (Rankin & Kern, 1994; Loeber & Loeber, 1986; Hoeve et al., 2012). The results of a recent study using nationally representative samples (Rankin & Kern, 1994), revealed that the strong attachment to both parents significantly decreased the probability of delinquent behaviors. Several meta-analyses have also found that parental attachment and parent-child involvement are strong predictors of problematic behaviors and delinquency (Loeber & Loeber, 1986; Hoeve et al., 2012).

The second domain is related to the social learning aspects. As argued by Akers (1992) the attitudes and beliefs towards crime and problem behaviors of individuals are learned by direct or indirect experiences. Moreover, the theory insists that motivation and criminally oriented beliefs are also learned and reinforced by interacting with intimate others. Because the closest role models of juveniles are their parents, it is natural to assume that the belief, techniques, and motivation of crime and deviant behaviors can be passed to their children by learning mechanisms. Empirical studies also showed that the parental criminality and antisocial life patterns of parents increased the likelihood of delinquency, while the effect is marginal (Laub & Sampson, 1988).

The third domain is associated with general strain theory (GST) argued by Agnew (1992). From the perspective of GST, stress-inducing experiences within the family such as parental rejection, parental incarceration, poverty, and abusive parenting skills yield negative emotions, in turn, individuals cope with the stress by engaging in antisocial behaviors. Empirical studies frequently examined the influence of family strain on delinquency and found that parental rejection (see Loeber & Loeber, 1986), stressful family relationships, and bad parenting habits (Agnew, Brezina, Wright, & Cullen, 2002) significantly increased the probability of delinquency and problem behaviors. For instance, Agnew and his colleagues (2002) examined the impact of strain generated from familial characteristics and parenting habits while controlling for the effect of parental attachment using nationally representative samples. Results of their study showed that family strain, parenting habits, and divorced/separated parents are significantly associated with delinquency, while parental attachment had no significant impact.

Peer and Environmental Factors

Association with deviant peers significantly increases the likelihood of criminal involvement and delinquency. According to differential association and social learning theories, children learn all behaviors including criminal behaviors through interactions with others (Sutherland, 1939). Especially, a relationship with peers is one of the most intimate social interactions for children, thus, they largely build their definition of good or bad norms via peer relationships, and the definition is also reinforced by observing the benefit and costs of peers' criminal behaviors (Burgess & Akers, 1947; Akers, 1998). Indeed, in several empirical studies, peer delinquency and attitudes toward the antisocial behaviors of peers are found to be strong predictors of delinquency and antisocial behaviors (see Pratt et al., 2010). In addition to the learning process, negative emotions emanating from bad experiences with peers may also increase the risk for offending as argued by GST. Along with direct victimization experienced by peers (Oh & Connolly, 2019), witnessing victimization of peers may also lead to delinquency (Kort-Butler, 2010). For instance, individuals who are being bullied by peers are more likely to show behavioral problems at a later stage of life and the association was partially mediated by increased negative emotions (Oh & Connolly, 2019).

Across the life of individuals, the school environment has a strong influence on their career and life standards. As argued by Hirschi's (1969) social control theory, crime and delinquent behaviors are more likely to occur when the social bonds are weak. According to the social control theory of Hirschi, individuals who are more attached to, committed to, and involved in school activity are refrain from offending crime. Empirical evidence on bonds to school also supports the argument (Wiatrowski, Griswold, & Roberts, 1981; Booth, Farrell, & Varano, 2008). Wiatrowski and his colleagues (1981) examined the influence of bonds to school on antisocial belief and delinquency using juvenile samples. The findings of their study showed that school attachment has a substantial influence on delinquency, while school commitment and involvement have relatively small effects on delinquency. Further, individuals who are devoted to school activity and received higher grades are more likely to have stable jobs in their adulthood, thus the opportunity for offending crime drops. Therefore, academic performance, bond to school, and school satisfaction at a younger age are appeared to be related to offending at a later stage of life.

In addition to the school environment, the neighborhood environment also affects the life-course trajectory of behavioral problems during childhood. Frequent exposure to violence and chronic disadvantages of neighborhoods are found to have negative effects on violent offending and delinquency. From the perspective of GST, individuals who are surrounded by disorganized neighborhood environments may experience higher levels of emotional strain, which in turn increases the likelihood of problem behaviors. In the study of Agnew and his colleagues (2002), neighborhood strain measured by asking respondents with "How is your neighborhood as a place for kids to grow up?" is significantly associated with delinquency while controlling for other individual-level risk factors.

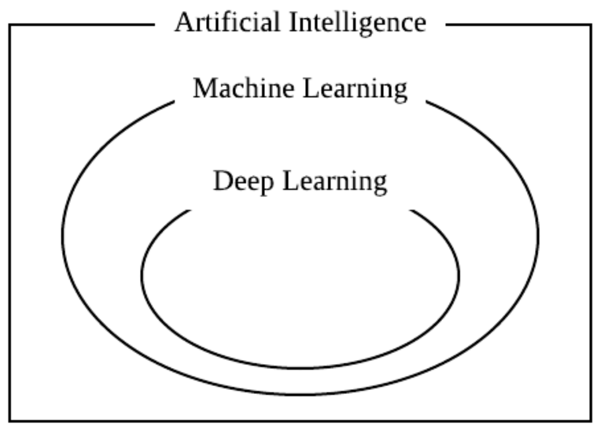
Machine Learning and Predictive Modeling

What is Machine Learning?

Machine learning, also known as statistical learning, is the statistical model that allows computers to perform a specific task by learning patterns and inferences using training data without being explicitly programmed (Berk, 2008; Samuel, 1959). ML is also one of the data mining approaches that are designed to build the best accurate mathematical model based on training data in forecasting outcomes. This computational algorithm is also considered as a subset of artificial intelligence, which is designed to form abstractions and concepts, or solve problems as can be seen in Figure 1 (McCarthy et al., 1955).

Figure 1

Artificial Intelligence, Machine Learning, and Deep Learning



Nowadays, it is widely acknowledged that artificial intelligence outperforms the human brain in learning and solving problems of games such as chess and GO games. The job of artificial intelligence was not limited to the areas of the game but expanded to almost all fields of science. The subfield of AI, machine learning, was applied to detect and diagnose disease (Sajda, 2006), to generate and test theories of molecular and material science (Butler et al., 2018), to predict the stock market (Patel et al., 2015), or to analyze and annotate a wide variety of genomic sequence elements (Libbrecht & Noble, 2015). Considering that the reasoning method of ML is inductive, the technique generally has been used to find the hidden patterns and relationships between observed cases. Stated differently, ML is more appropriate to create hypotheses and forecast the outcome based on self-learned algorithms than explaining causal relationships based on existing theories (Hofman et al., 2017). Because of the inductive reasoning feature, ML became vigorously used in the field of hard science in excavating more complex theories and predicting the outcome based on ML-generated hypotheses.

This section of the literature review presents specific types of ML techniques that are applicable to predicting antisocial behaviors of individuals using risk assessment instruments. In the first part of this chapter, a brief overview of the types of Machine Learning is illustrated. There are three types of machine learning based on supervision types; supervised learning, unsupervised learning, and reinforcement learning. The type of learning process is decided by whether or not the supervisor of the learning process exists, and by how the supervisor evaluates the learning process. Among three types of ML approaches, the current study only discusses supervised learning because the learning process will be supervised by output variable, LCP or AL. Several methods of supervised learning are discussed further below, then, empirical studies predicting antisocial behaviors and recidivism based on ML are summarized in this part.

Supervised Learning

Supervised learning is the most common machine learning method when the purpose of the ML is to predict the output. In terms of supervised learning algorithms, the computer generates the program by questioning and answering based on training data which is comprised of paired samples of questions and answers (Sugiyama, 2015).

$$\text{Paired Sample: } \{(\mathrm{xi}, \mathrm{yi})\}_{i=1}^{n}$$

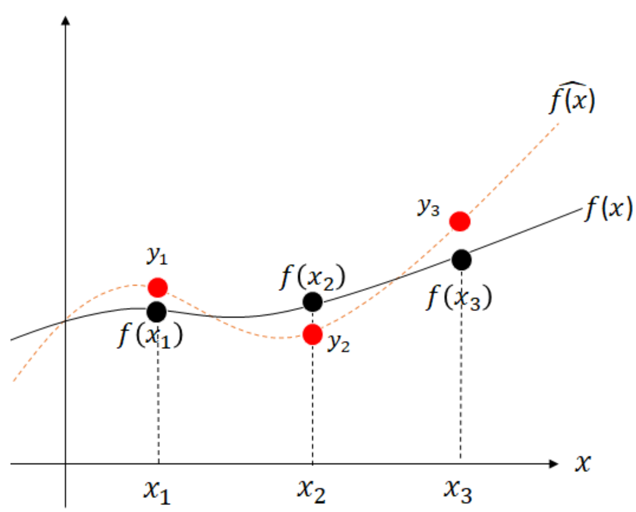
Where, x is input and y is output

The term *supervise* refers that there is training data with correct answer keys. The role of the computer in the ML process is similar to the role of students in the classroom. The computer keeps submitting answers repeatedly, then the answers are evaluated by the supervisor. Through the process of repeated questioning and answering, the computer creates the map to find routes from input to correct output. Depending on the types of output, supervised learning can be subcategorized into regression and classification.

Regression. When the output y is continuous, the regression line can be generated by a supervised learning process. The learning process is targeted to generate a function *f* corresponding to the supervisor's knowledge (Sugiyama, 2015). Similar to the general logic of linear regression analyses, the regression line which creates the least summed error can be generated from the learning process.

Figure 2

Example of Regression using Machine Learning



To minimize the size of the error, the regressed line can be linear or polynomial. By repeating the testing process using data unused for the learning process, the generalizability of function *f* can be evaluated. The testing data is the subsample of the total pairs of x and y, and should not be used during the learning process. The degree of generalization can be measured by the distance between the true function *f* and its approximation function.

$$Y = \theta_1 + \theta_2 x$$

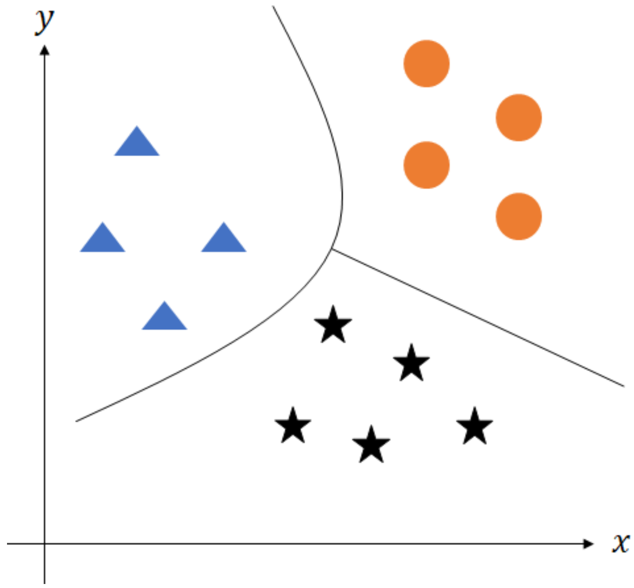
When regression analysis is used in machine learning, the regression line is not static but may be modified depending on the feedback from the testing process. The conventional regression analysis is designed to generate a line that most fits the whole data. When it comes to the Ordinary Least Square regression, the regression line is found by the principle of least squares in that the linear function minimizes the sum of the squares of the differences between the observed value and predicted value by the function. While conventional regression analysis can generate a single linear function, ML-based regression analysis builds an initial linear function and modify the function that minimizes the Root Mean Squared Error (RMSE) by the iterative testing process. The coefficients of the function in ML are calculated using a Gradient Descent approach. In the Gradient Descent process, the machine randomly selects the values of coefficients, and calculates the gradient of RMSE on selected coefficients, then updates the coefficients until the model finds a minimum value of RMSE (Sugiyama, 2015).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y_i}\right)^2}$$

Classification. When the output *y* is categorical, the dimensional vector *x* can be classified by a categorical scalar *y* {1,2,3,…c}. The paired samples of x and y are used when training the classifier like other supervised learning techniques. Classification is widely used when predicting the dichotomous output (i.e. yes or no) or categorical output without orders. The most frequently used classification algorithms are Tree-based classification methods, Neural Networks, Support vector model, and Discriminant Analysis.

Figure 3

Example of Classification



Measures of Performance

Conventional explanatory modeling of research is primarily targeting to examine the covariation between independent variables and dependent variables. One of the most frequently used statistical analyses, ordinary least squares, for instance, generates an imaginary straight regressed line most fitted to the observed data. The most fitted regressed line is generally determined by the least-squares approach, which generates the least sum of the squares of the differences between the observed value of the dependent variable and those predicted by the regressed line. When examining the association between independent variables and the dependent variable, the statistical significance of the coefficients is assessed within certain levels of confidence. The overall performance of the linear function can be assessed by computing the *R*-squared value. R^2 value represents the proportion of the variance for a dependent variable that is explained by an independent variable or variables in a regression model. Another performance parameter in regression analysis is Root Mean Squared Error (RMSE). RMSE is a summary of statistics for the overall difference between the observed and predicted value of the dependent variable. Lower RMSE represents the better performance of prediction.

In the predictive modeling approach, the overall performance of models is commonly measured by the classification error or complement accuracy. The main purpose of ML models is to predict the outcome using input data. Accordingly, the predictive parameters can be computed by assessing the proportion of successful prediction to failed prediction. Based on the test results, five measures can be calculated to show the predictive performance from the confusion table.

Table 1

Example of Confusion Table

		Observed Value	
		0	1
Predicted Value	0	True Negative	False Negative
	1	False Positive	True Positive

Sensitivity, also called the true positive rate, refers to the proportion of true positives among those who are observed as positive. For instance, the percentage of actual recidivists who are correctly predicted as recidivists by the predictive model is sensitivity. Specificity, also called the true negative rate, refers to the proportion of true negatives that are correctly identified as such. For instance, the percentage of non-recidivists who are correctly identified as non-recidivists by the predictive model is specificity. Accuracy refers to the percentage of correct assessment. Since the trade-off association between sensitivity and specificity, each measure alone does not show the performance of the predictive model in general. Precision refers to the proportion of true positives among those who are predicted as positives. F1-score, also called Harmonic Mean, is a measure combining precision and recall score in order to measure the balance between precision and recall.

Sensitivity = True positive / (False positive + False negative)

Specificity (Recall) = True negatives / (True negatives + False positives)

Accuracy = (True positive + True negatives) / (True positive + False positive + True negative + False negative)

Precision = True positive / (True positive + False positive)

*F1-score = 2 * (Precision*Recall) / (Precision + Recall)*

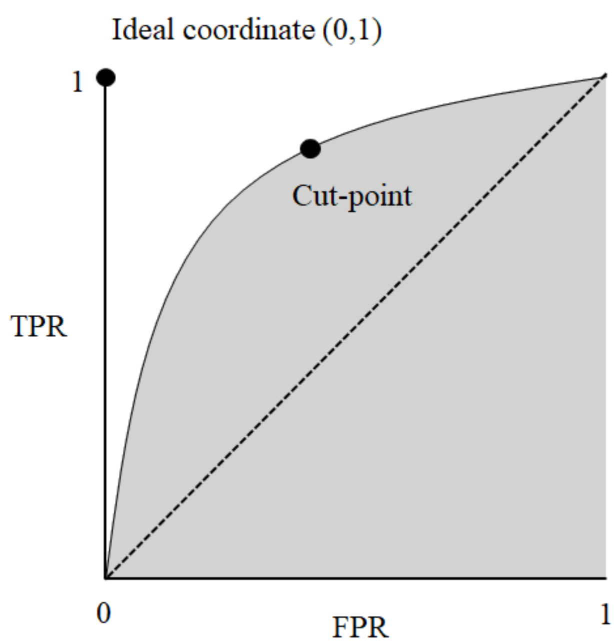
One of the most commonly used parameters of predictive performance is evaluated with receiver operating characteristics (ROC) analysis (Mossman, 1994). For a predictive model, the association between the true positive rate (TPR) against the false positive rate (FPR) can be drawn as Figure 4. The curve, called as Receiver Operating Characteristics (ROC) curve, indicates the amount of false-positive results against the amount of false-negative results. As can be seen in Figure 4, a certain point can be determined to increase true positive rates while minimizing false-positive rates. However, depending on the focus of the test, the researcher can select the cut-off points of the test to increase or decrease the TPR and FPR. Therefore, the areas under the curve (AUC) in ROC is calculated to determine relative accuracy for the predictive model according to variant FPR ranging from 0 to 1. For instance, AUC 0.5 tells that the model is not valid for the classification because the chance of getting true positive against false positive is as same as the chance of flipping coins (random chance). Higher AUC indicates that the model performs better in predicting correctly. In general, predictive models with a higher than .70 of AUC are considered acceptable (Hosmer & Lemeshow, 2000).

TPR = True positive / (True positive + False negative)

FPR = False positive / (False positive + True negative)

Figure 4

ROC Space: Shadow Area Represents the Area under Curve



Application of ML in Predicting Antisocial Behaviors

Despite the increasing attention to artificial intelligence in almost all fields of science, there are only a few empirical studies have attempted to apply machine learning algorithms in predicting criminal or antisocial behaviors. Moreover, fewer empirical studies examined the practical implication of policies utilizing machine learning-based predictive tools. Nevertheless, a series of empirical studies have been vigorously conducted by scholars at the University of Pennsylvania. In addition to these scholars, a small number of scholars delved into the application of ML methods in the criminal justice system and criminology studies. In this part, empirical studies using ML algorithms in predicting recidivism and other behavioral problems are presented. Firstly, studies using the tree-based classification model, are discussed in the first segment. Since the model has been emerged and developed in the early stage of artificial intelligence technology, a relatively large amount of study can be discussed in the first section. Next, empirical studies applying neural networks to predictive models for behavioral problems are discussed.

Predicting Antisocial Behavior using Tree-Based Model

Gardner and his colleagues purported that the actuarial prediction methods of violence for clinical purposes were impractical since the prediction methods based on regression analysis of risk factors were difficult to apply in clinical settings (Gardner et al., 1996). Moreover, they also argued that predicting a small number of violent patients out of the general psychiatric population is too costly. To overcome the obstacles, Gardner and his colleagues suggested applying the regression tree and two-stage screening methods to predict violent patients. The data was drawn from a previous study

(Lidz, Mulvey, & Gardner, 1993) that interviewed 784 patients who visited emergency rooms. Through the follow-up community interview of the patients, the violence of the subjects was measured by records of arrest or by asking about their involvement in violent incidents. The study generated the regression trees (CART) to forecast the violence after discharge using predicting variables including individuals' sociodemographic characteristics, use of substances, and the degrees of distress measured by the Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983). The results of the predictive model built on CART were compared with the results of negative binomial regression (NBR) by comparing AUC, sensitivity, and specificity. Their findings revealed that both predictive models, using CART and NBR, produced indifferent performances in predicting violence.

Mental health researchers continued to apply the tree-based predictive model to forecasting violent behaviors and mental disorders. Steadman et al. (2000) examined the predictive ability of the classification tree model, Chi-squared Automatic Interaction Detector (CHAID), using 939 patients discharged from acute psychiatric units of hospitals. According to the existing theories and findings from empirical researches, the study identified 134 risk factors of violence and measured each item during the subjects' hospitalization. The study generated three predictive models, one-stage CHAID, iterative CHAID, and Logistic regression model. Findings of the analysis revealed that the predictive ability of the iterative CHAID model (AUC = .82) is slightly better than that of logistic regression (AUC=.81), whereas the one-stage CHAID model performed worst (AUC = .79) among the three models.

Applying the iterative classification procedure of Steadman et al.'s (2000) study, Silver, Smith, and Banks (2000) attempted to forecast recidivism. The study focused on comparing the predictive performances of the iterative classification models to those of traditional screening devices, such as Burgess (unweighted summation of risk factor scores), logistic regression, and CART. Using 11,714 convicted offenders' information drawn from the New Jersey Administrative Office of the Courts, Silver et al. (2000) constructed five predictive models with 14 risk factors that are measured by official records of offenders. Four performance parameters of each model, Mean Cost Rating (MCR), Gini Index of Diversity, Dispersion Index for Risk, and % classified as a high or low-risk group were calculated for the model evaluation. MCR represents the predictive accuracy which is a scaled version of the AUC (MCR = (AUC - .5/.5). However, findings revealed that when comparing MCR between models, the standard CART and iterative CART model did not outperform conventional screening methods such as Burgess, and LR. In specific, the MCR of CART was the lowest among those of the five models in all cases. Although the iterative CART model showed better predictive effectiveness than the standard CART model, it was still less effective than logistic regression.

Rosenfeld and Lewis (2005) also applied the CART model to predict violence among stalking offenders using risk factors. The data used in the study was drawn from the official records of criminal defendants referred to the New York City Forensic Psychiatry Clinic between 1994 and 1998. Using 204 individual offender samples who are convicted of stalking or obsessional harassment incidents, three predictive models were developed by expanding the number of predictors for each successive model. The first model includes five risk factors age, education, the threatened victim, prior intimate relationship, and revenge motivation. The second model expands the risk factors by adding psychotic disorder, personality disorder, substance abuse history, and criminal history. The third model includes prior violence, below-average intelligence, gender, and foreign-born along with all predictors used in the second model. Results of analyses revealed that the CART model yielded better predictive performance than the logistic regression model with the higher AUC. In specific, the full model with all predictors yielded the highest area under the curve (AUC) of .848, with significant improvement over the logistic regression model (.801).

Table 2

Predictive Accuracy in Rosenfeld and Lewis (2005)'s Study

	Model 1		Model 2		Model 3	
	AUC	90% CI	AUC	90% CI	AUC	90% CI
Tree regression	.787	.73-.84	.836	.79-.88	.848	.80-.89
Logistic regression	.780	.73-.83	.800	.75-.85	.801	.75-.85
Tree-cross-validation	.659	.59-.73	.644	.58-.71	.649	.58-.72
Logistic cross-validation	.744	.69-.80	.725	.66-.79	.706	.64-.77

Note. Reprinted from Roselfeld and Lewis (2005)

Berk, Li, and Hickman (2005) attempted to utilize CART and Random Forest models in predicting probation and sentencing decisions from a different perspective than previous studies. The study focused on the effects of the race of defendants and victims on capital charges and death penalties. Contrasting to previous studies using ML, the purpose of the study was to compare the role of race in capital cases when applying tree-based predictive models and when applying logistic regression models. To examine the impact of one variable, the variable was replaced with the random values (0s and 1s) and the changes in overall prediction errors between the shuffled model (with random numbers in a substituted A) and the original model (with a real variable A) was evaluated. For the analysis, 1,311 death-eligible cases were drawn from the previous study which reported that the cases with white victims and black defendants are more likely to be sentenced to be charged with a capital crime and sentenced to death (Paternoster et al., 2003). In contrast to the findings of the previous study using conventional parametric statistical analysis, Berk, Li, and Hickman (2005) found that the race of defendants and victims does not show significant impacts on prediction error for capital charges and death sentences. As can be seen in Table 3 and 4, the prediction error increases by approximately 20 percent when shuffling race variables, while the prediction error considerably increases when shuffling the County variable (Baltimore City) or numbers of victims. As concluded by Berk and his colleagues (2005), conventional statistical modeling was found to be very unstable to specify the role of certain variables in predicting outcomes. Moreover, CART and Random Forests models yield significantly different results than logistic regression models regarding the impact of race on capital charges and death sentences.

Table 3

Prediction Error for Cases Charged with a Capital Crime in Berk, Li, and Heckman (2004)'s Study

Predictor Shuffled	Prediction error
None	.19
White defendant-White victim	**.20**

(continued)

Black defendant – Black victim	**.19**
Other race	.23
Baltimore City	**.33**
Previous felony	.19
More than one victim	.21

Note. Reprinted from Berk, Li, and Heckman (2004).

Table 4

Prediction Error for Death Sentence Cases in Berk, Li, and Heckman (2004)'s Study

Predictor Shuffled	Prediction error
None	.20
White defendant-White victim	**.22**
Black defendant – Black victim	**.21**
Other race	.23
Baltimore City	.25
Previous felony	.25
More than one victim	**.30**

Note. Reprinted from Berk, Li, and Heckman (2004).

Berk, Kriegler, and Baek (2005) applied CART and Random Forest to forecasting inmates' misconduct. Using 9,662 male inmate samples drawn from the California Department of Correction data, inmates' serious misconducts while incarceration was predicted by two ML methods, CART and RF, and logistic regression. Berk et al. (2005) argued that the predictive accuracy is not meaningful if the outcome variable is highly imbalanced because the model can yield very low classification error rates by classifying all cases to negative or positive classes. For instance, when 1,000 random samples were selected from the data, there were only 39 (3.9 %) inmates engaged in serious misconduct out of 1,000 cases. In this case, the baseline predictive accuracy of classification is 96.1 % by classifying all inmates to the no misconduct class. The study pointed out the pitfall of predictive accuracy in imbalanced data and suggested the manipulation of the cost ratio. Instead of the default cost ratio of 1.0, the study intentionally increased the cost of false negatives to false positives. The authors determined that misclassifying false negative, who would actually engage in serious misconduct but predicted as non-misconduct, is more costly than false positive, who would not actually engage in serious misconduct but predicted as misconduct. As can be seen in Table 5, 51.3 % of actual misconduct was correctly predicted by RF with a 10 to 1 cost ratio assuming that the cost of missing actual misconduct is ten times higher than a false forecast of misconduct. In a similar way, 38.5% of actual misconduct was correctly predicted by the RF model with a 5 to 1 cost ratio (Table 6). Moreover, the importance of each predictor in predicting inmates' misconduct was evaluated by shuffling each predictor. The variable importance plot, Figure 5, revealed that the term length yields the most contribution to forecasting serious misconducts accurately. Overall, the study of Berk, Kriegler, and Beak (2005) provided the alternative application of the Random Forest model in predicting behavioral problems by manipulating cost ratios of false negatives to false positives.

Table 5

Random Forest Confusion Table with 10 to 1 Costs in Berk, Kriegler, and Baek (2005)'s Study

	Forecast no misconduct	Forecast misconduct	Model error
Observed no misconduct	753	208	.216
Observed misconduct	19	20	.487
Use error	.024	.912	Overall = .227

Note. Reprinted from Berk, Kriegler, and Baek (2005).

Table 6

Random Forest Confusion Table with 5 to 1 Cost in Berk, Kriegler, and Baek (2005)'s Study

	Forecast no misconduct	Forecast misconduct	Model error
Observed no misconduct	837	124	.129
Observed misconduct	24	15	.615
Use error	.028	.892	Overall = .148

Note. Reprinted from Berk, Kriegler, and Baek (2005).

Figure 5

Average Reduction in Forecasting Skill for Serious Misconduct in Berk, Kriegler, and Baek (2005)'s Study

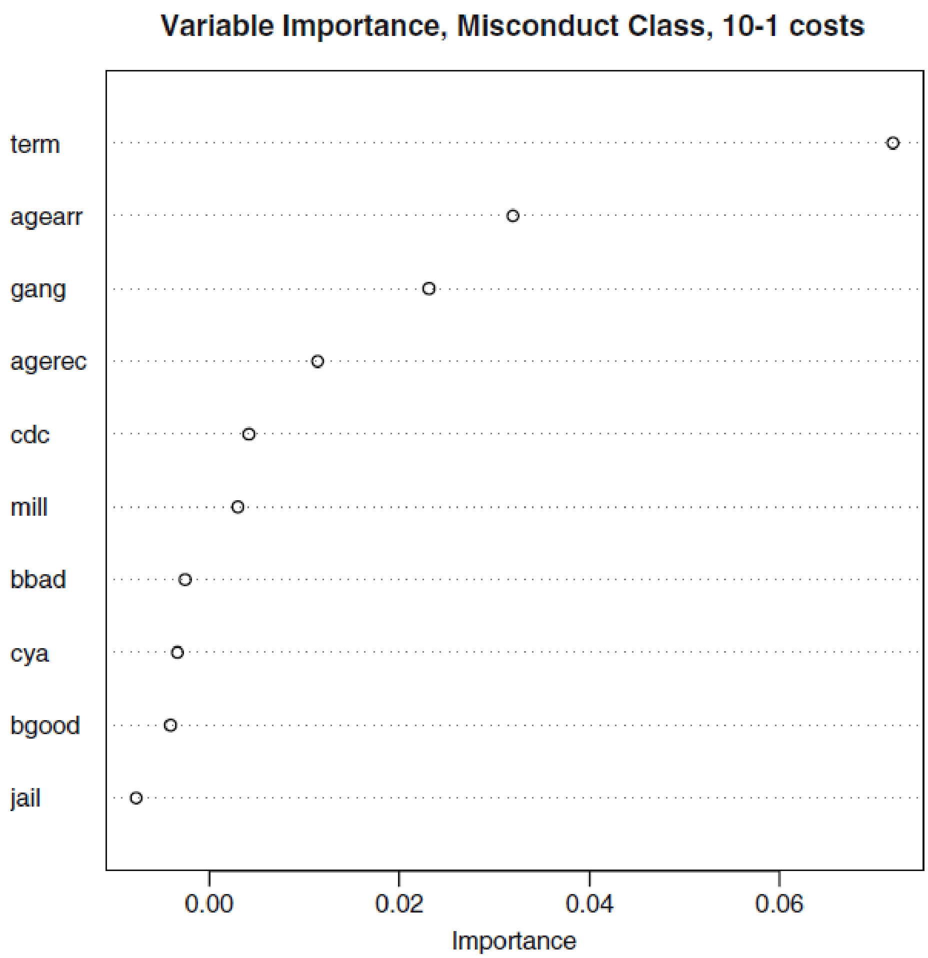


Fig. 1 Average reduction in forecasting skill for serious misconduct

Note. Reprinted from Berk, Kriegler, and Baek (2005), p. 139.

Berk applied the same processes of his previous study (Berk, Kriegler, & Baek, 2005) to Philadelphia's Adult Probation and Parole Department (APPD) data. Berk, Sherman, Barnes, Kurtz, and Ahlman (2009) constructed RF models to forecast murder with randomly selected 30,000 cases drawn from probation and parole data of APPD collected between 2002 and 2004. With the 10 to 1 cost ratio settings, the RF model correctly predicted 43% of homicide offenders within the parolee and people on probation samples, whereas the logistic regression model misclassified 99.7% of homicide offenders. The variable importance evaluation revealed that age is the most significant predictor of homicide, followed by age of first contact, prior gun violations, and gender. In addition to the variable importance, the study calculated how age is related to the odds of being charged with a homicide or an attempted homicide. The partial dependence plot shows that the association between age and log-odds of being a homicide offender is non-linear, furthermore, the log odds sharply decrease after age 30 years. It is concluded that predicting recidivism with the RF model may benefit practitioners by providing a piece of fruitful information to target limited resources to high-risk parolees and people on probation.

Another attempt has been made to predict recidivism among mentally disordered offenders in the field of psychiatric studies. Pflueger et al. (2015) aimed to build the predictive model of recidivism using 259 mentally ill offenders in Switzerland with an individual's criminal records, sociodemographic information, and clinical diagnoses regarding the mental disorder. Unlike previous studies, Pflueger et al. (2015) applied the random survival forests model which allows the modeling of right-censored survival data (see Ishiwaran, Kogalur, Blackstone, & Lauer, 2008). Results of the final model using random survival forests modeling identified that the most important predictor is the number of prior convictions followed by age, type of index offense, diversity of criminal history, and substance abuse. Interestingly, clinical diagnoses were not found to be important in predicting criminal recidivism except for the history of substance abuse.

Although the predictive performance of the final predictive model built on the random survival forest modeling technique was not compared to conventional statistical analysis, the performance parameter of their predictive model was outstanding (AUC = .90).

Figure 6

Predictor Importance for Forecasting Skill in Berk, Sherman, Barnes, Kurtz, and Ahlman (2009)'s Study

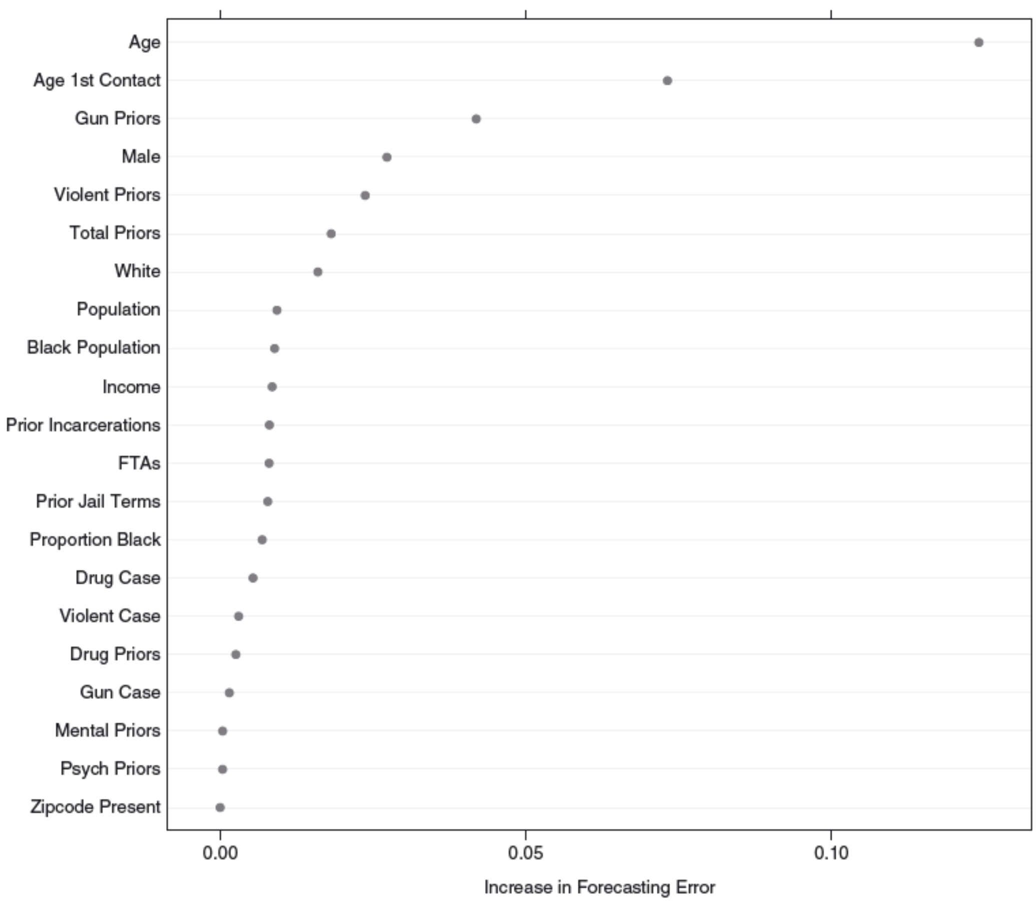


Fig. 1. Predictor importance for forecasting skill

Note. Reprinted from Berk, Sherman, Barnes, Kurtz, and Ahlman (2009), p. 200.

Table 7

Random Forests Confusion Table for Forecasts of Homicide or Attempted Homicide by using the Training Sample and out of the Bag Observations in Berk, Sherman, Barnes, Kurtz, and Ahlman (2009)'s Study

	Forecast no homicide	Forecast homicide	Model error
Observed no homicide	27914	1764	.06
Observed homicide	185	137	.57
Use error	.007	.93	Overall = .07

Note. Reprinted from Berk, Sherman, Barnes, Kurtz, and Ahlam (2009).

Neuilly et al. (2011) applied random forest models to forecasting recidivism of released homicide offenders using 320 randomly selected homicide offenders released between 1990 and 2000 in the New Jersey Department of Corrections. The study constructed two predictive models based on the random forests model and logistic regression-based classification tree model. Within the five years after release, recidivism was predicted with a wide range of predictors, such as demographics, types of charged offenses, prior criminal history, and lifestyle characteristics of individuals. Neuilly et al (2011) found that the predictive performance of the random forest model outperforms those of the logistic regression-based classification tree model.

Berk and Bleich (2014) attempted to apply the classification tree model of ML on predicting recidivism. The Philadelphia probation data between 2002 and 2005 were used to train the classifier model and test the performance of the model. Using 48,923 individual cases with pairs of input and output, the computer learned the patterns and generated classification trees. In training data, the output is coded as two categories, 1 for recidivism and 0 for non-recidivism. Input predictors were demographic information (gender, date of birth), juvenile prior records (total number of priors, number of sex offense priors, number of drug priors, etc), adult prior records (total number of priors, number of sex offense priors, number of murder priors, etc), and other criminal records information.

The performance of two different types of classification trees was compared. As a benchmark, the random forest classifier is generated because the authors believed that the model performs best for criminal justice behavioral forecasting (Berk, 2012). The performance of the classification and regression tree (CART) and the random forests model were generated by a supervised learning process using training data. The remaining subsample is not used in the learning process and is only used for the evaluation of the predicting performance of the learned classifier.

In Berk and Bleich's (2014) study, the cost ratio is adjusted for the sake of increasing the chances of false-positive while decreasing false-negative. In practice, the decision-makers and researchers should determine the relative costs between false-negative and false-positive when making predictions. Berk and Bleich (2013) adjusted the cost ratio so that false-negative become 5 times more costly than false positives. Stated differently, they believe that failing to predict high-risk individuals is 5 times more harmful than incorrectly predict high-risk individuals. Berk (2012) argues that asymmetric costs and tuning procedures are dependent on stakeholder's preferences.

As a conclusion of their prediction, a simple symmetric classification tree 60% correctly predicted individuals who fail and who do not fail. After adjusting the cost ratio between false negative and false positive, the forecasting model accuracy for individuals who do not fail is 92%. However, as the countering effect of high accuracy of predicting “no-fail”, the asymmetric forecasting model predicted “fail” only 21% correctly.

Table 8

Random Forests Confusion Table using 5 to 1 Cost Ratio of False Negatives to False Positives in Berk and Bleich (2014)'s Study

	Forecast Fail	Forecast No Fail	Model error
Observed Fail	3,331	1,697	.34
Observed No Fail	8,837	25,100	.26
Use error	.73	.06	Overall = .27

Note. Reprinted from Berk and Bleich (2014).

Table 9

Classification Tree Confusion Table using Test Data in Berk and Bleich (2014)'s Study

	Forecast Fail	Forecast No Fail	Model error
Observed Fail	719	515	.42
Observed No Fail	2,829	5,722	.33
Use error	.80	.08	Overall = .342

Note. Reprinted from Berk and Bleich (2014).

There was a substantive amount of concern about the imbalance of outcome value when applying ML algorithms to predicting recidivism or criminal behaviors (Chen, Liaw, & Breiman, 2004; Berk, 2008). To remedy the potential problems due to the imbalanced outcome value, two applied random forest models can be adopted: weighted

RF and balanced RF (Holleran and Stout, 2017). Holleran and Stout (2017) applied the balanced RF to forecasting the outcome of family court according to its superior results in previous empirical studies. Using 19,326 cases of New Jersey Family Court, the outcome of each case including dismissed, noncommitment, and commitment were classified by the balanced RF model. Instead of focusing on overall accuracy, the study majorly focused on minimizing the misclassification error for commitments. As argued by Berk (2008), the costs of misclassifying actual risky subjects are higher than those of misclassifying non-risk offenders as risky. Results of the study indicated that the balanced RF shows significant decreases in the misclassification error for commitments. In specific, the balanced RF misclassified 7% of actual commitment cases as non-commitment, whereas balanced RF misclassified 78% of actual commitment cases as non-commitment cases. Although the overall accuracy of the balanced model (19.6%) is lower than those of an balanced model (43.6%), Holleran and Stout (2017) argued that class-specific error should be more paid attention to than overall accuracy when considering the class imbalance on the dependent variable. Furthermore, the study also examined the role of the race of juveniles on the decision of courts by dropping the race variable during the statistical learning procedure. The results revealed that the race of juveniles rarely contributes to the commitment classification of the model.

Predicting Antisocial Behavior using Neural Networks

The first empirical study applying Neural Networks to recidivism prediction was done by Caulkins, Cohen, Gorr, and Wei (1996). Caulkins and his colleagues attempted to apply NNs to investigate additional information from Gottfredson and Gottfredson's (1979, 1980, 1985) federal prison inmate data. The data is comprised of 29 predicting variables regarding offense type, criminal history, social history, and institutional adjustment, of 3,508 ex-offenders released from federal prisons in 1970 and 1972. Of those released offenders in 1970 and 1972, 1,207 releasees (34.41 %) were reported to recidivists. The NNs learned the predicting model using data of released offenders in 1970 and test the predictability of the learned model using the data of 1972. In specific, the machine is trained through a dataset drawn from 1970 and established the predictive model, in turn, the predictive model was applied to forecast the recidivism of released offenders using 1972 data. Although it is unusual to use two different datasets for training and testing processes instead of randomly splitting a dataset into training and testing data, Caulkins et al. (1996) were able to use different datasets for each process since both datasets include common predicting factors and response variable.

To compare the predictive accuracy of recidivism between the multivariate regression model and the NNs model, Caulkins et al. (1996) replicated the regression analysis which is originally run by Gottfredson and Gottfredson (1979). Among several different types of NNs models, the authors selected a multilayer neural networks model, and the quick propagation algorithm was adopted to train the model. Five performance parameters for both models, multivariate regression and NNs, were calculated to evaluate whether NNs is outperformed in predicting recidivism than multivariate regression model: False-positive rate (FPR), false-negative rate (FNR), the percentage of total correct predictions (TCP), a relative improvement over chance (RIOC), and mean cost rating (MCR). Except for the first two parameters, false-positive rate and false-negative rate, values close to 1 indicate greater predictive accuracy. The results of the analysis revealed that all parameters of the NNs model are better than its counterpart when predicting recidivism in the training data drawn from 1970 data. However, testing the performance of the model using training data yields overfitting issues. Therefore, the performance of the learned model should be calculated within testing data that were not used for the training. When it comes to the performance parameters calculated using testing data, almost all parameters of NNs were reported to be worse than the multivariate regression model. In summary, Caulkins et al (1996) concluded that neural network models do not outperform over multiple regression analyses in predicting recidivism. As can be seen in Table 10, MCRs of neural network models using 11 predicting variables calculated from testing data were lower than those of multiple regression. It is consistent with findings from previous studies predicting complex human behaviors using neural network models (Dawes, 1979).

Table 10

Comparison of Predictive Accuracy of Seven Methods in Caulkins et al. (1996)'s Study

Model	Construction Data MCR	Validation Data MCR
Burgess: 19 items	.408	.404
Multiple regression: 11 vars	.440	.436
Association analysis: 8 terminal groups	.338	.328
Predictive attribute analysis: 11 terminal groups	.429	.389
Multidimensional contingency table analysis	.419	.397
Multiple regression (replication): 11 vars	.431	.432
Neural Networks model: 11 vars	.460	.416

Note. Reprinted from Caulkins et al. (1996).

Palocsay, Wang, and Brookshire (2000) also attempted to build neural network models for criminal recidivism forecasts. Their study majorly focused on examining the performance of NNs and investigating the advantages of NNs over traditional statistical models. Using relatively large data of 19,136 released offenders from North Carolina prison between 1977 and 1980, the performance of the predictive model using NNs was compared to those of the predictive model using logistic regression. The data is comprised of two sets of data published in 1978 and 1980. The 1978 data contained information of 9,457 offenders released between July 1st, 1977 and June $30^{th,}$ 1978, while 1979 data contained information of 9,679 offenders released between July 1st, 1979 and June $30^{th,}$ 1980. After dropping incomplete records, a total of 10,617 cases of the released offender was used in the analysis.

Consistent with Caulkins et al.'s (1996) study, Palocsay and his colleagues (2000) adopted multi-layer neural networks modeling to construct the predictive models in the NeuroShell 2 software. Palocsay et al. (2000) improved their learning algorithms by applying backpropagation which is designed to adjust weights according to the network's inter-node connections that minimizes the total error functions. After running the training processes, they selected the 26-node network model that had the highest percentage of test-set correct classification with a smaller network configuration. Similar to the evaluation conducted by Caulkins et al. (1996), the trained NNs model with 26-node networks using 1978 data is applied to the 1980 test data to avoid overfitting issues. To evaluate the predictive accuracy of models, three parameters were calculated, Odds ratio, Yule's Q, and relative improvement over chance (RIOC). The odds ratios refer to the ratios of precision (TP / TP+FP) to the false omission rate (FN / FN + TN). Stated differently, the odds ratio was calculated by dividing the odds of being a recidivist for those who were predicted to be recidivists by the odds of being recidivists for those who were predicted to be non-recidivists. A higher odds ratio indicates a more successful prediction. Similarly, values closer to 1.00 of Yule's Q and RIOC are indicating a more successful prediction.

Table 11

Classification Accuracy for Test Data in Palocsay et al. (2000)'s Study

Model	Recidivist correct (%)	Non-recidivist correct (%)	Total correct (%)
1978 Neural Networks	41.36	85.89	69.23
1978 Logistic regression	30.41	88.43	66.73
1980 Neural Networks	40.93	82.63	66.98
1980 Logistic regression	30.53	86.84	65.71
1978/1980 Neural Networks	39.01	82.15	65.96
1978/1980 Logistic regression	36.35	81.07	64.29

Note: Reprinted from Palocsay et al. (2000).

Table 12

Measures of Association for Test Results in Palocsay et al. (2000)'s Study

Model	Odds ratio	Yule's Q	RIOC
1978 Neural Networks	4.291	.622	.419
1978 Logistic regression	3.325	.537	.377
1980 Neural Networks	3.297	.534	.337
1980 Logistic regression	2.898	.486	.331
1978/1980 Neural Networks	2.844	.492	.308

(continued)

1978/1980 Logistic regression	2.446	.419	.257

Note. Reprinted from Palocsay et al. (2000).

Although the total correct rates for NNs and LR are slightly different, the model parameters of NNS were significantly better than LR. According to McNemar's test (see Lachenbruch, 2014), the predictive accuracy of the NN model is significantly higher than that of the LR model when the model was trained in 1978 data and tested using 1980 data. Also, the examination of the effects of different cut-off values on classification accuracy revealed that the NN model is outperformed the LR model in predicting recidivists and non-recidivists at all degrees of cutoff value. Accordingly, the author concluded that the recidivism predictive model using NNs is more accurate than the model using LR.

The application of NNs to predicting recidivism has also been examined in Sweden (Grann & Langstrom, 2007). The main purpose of the study was different from those of other studies. Instead of examining the prediction accuracy of a machine-learned predictive model using NNs, the study focused on examining whether weighted risk factors based on the results of the NNs model predict recidivism better than using total risk scores based on non-weighted risk factors. Individuals included in the study were those who have been diagnosed with either personality disorder or schizophrenia between 1988 and 1993 in Sweden. The response variable, recidivism, was recorded if the individuals are convicted of a violent crime during two years of follow-up period after release or discharge. Grann and Langstrom (2007) selected 10 items out of the HCR-20 risk assessment tool and ran four different statistical models to generate a weight loading for each risk factor item. The four statistical models they used were the Nuffield procedure, bivariate logistic regression, multivariate logistic regression, and NNs. Based on the results of four statistical analyses, the weight loadings were applied to generate total risk scores. Since the sample size is relatively small (N=404), Grann and Langstrom (2007) used the area under the curve (AUC) as a parameter of predictive validity. However, the results revealed that AUC was the highest when total scores of non-weighted risk factors were used to predict violent recidivism, whereas the weighted risk factor model based on the results of the Neural Networks model shows the lowest AUC. Therefore, the authors concluded that applying weights to the risk factors does not improve the accuracy of predicting recidivism. Conversely, the more complex algorithm for weighting, the worse the predictive accuracy was found.

Despite the undesirable findings of Grann and Langstrom (2007), the results should not be interpreted as evidence of the null effect of NNs models in predicting recidivism. Since the purpose of the study was to improve risk assessment tools using coefficient estimated by NNs models, the authors independently applied the NNs model for the sake of estimating the relative importance of each risk factor. Stated differently, the NNs model applied in this study was not designed and built to forecast the recidivism, instead, the NNs model was only used to calculate the weights of each risk factor. Therefore, the performance parameter AUC was calculated by predicting recidivism based on conventional methods, total scores of risk assessment tools with or without weights for each risk factor.

In 2010 the Ministry of Justice in the UK reported the results of NNs models in predicting serious offenders and recidivism (Yang, Liu, and Coid, 2010). While, Yang and his colleagues (2010) acknowledged inconsistent findings of NNs and other machine learning-based statistical models in predicting recidivism, the applicability of NNs and the classification and tree models in predicting violent recidivism and interactions between different elements of samples were aimed to be examined. Using 1,353 adult male and 304 female samples drawn from the Prisoner Cohort Study samples which are comprised of released offenders between 14th Nov 2002 and 7th Oct 2005 from the prison of the UK (see Coid et al., 2007). During the 4 years of follow-up interview, 45 percent of male releasees were reconvicted, while 28.9 percent of female releasees were reconvicted.

Four types of psychometric risk assessment instruments were used to identify the predicting variables including HCR-20, PCL-R, VRAG, and RM2000V. Then four models using different sets of predictors were trained for each machine learning algorithm. Violence Risk Appraisal Guide (VRAG) includes 12 predictors, Psychopathy Checklist-Revised (PCL-R) includes 20 predictors, Historical-Clinical Risk Management 20 (HCR-20) includes 20 predictors, while Risk Matrix 2000-Violence (RM2000V) is only comprised of 3 predictors. To predict violent recidivism, the authors applied four types of statistical analysis including logistic regression, discriminant analysis, classification tree, and multi-layer perceptron neural networks. In total, 16 models of statistical analysis were possibly built in the study. To compare the predictive performance between models, the study calculated the sensitivity, specificity, and overall accuracy.

The study generated the predictive models for two different outcomes, reconviction and violent recidivism. Focusing on the overall accuracy of test data, the NNs model outperformed other predictive models for the male samples. On the contrary, NNs models did not show superior performance when it comes to models for the female sample. Although the overall accuracy of NNs models using test data outperformed other models within the same sample and instrument tools, sensitivity dramatically in almost all cases (.16-.53). In summary, the authors concluded that the performance of machine learning-based models demonstrated no significant improvement in predicting recidivism and violent re-offense, regardless of the type of applied predicting variables.

Despite the nonsignificant improvement of the NNs model in predicting recidivism and violent offenses over conventional statistical methods, the authors advanced their predicting model by fine-tuning the predicting and outcome variables. Only the NNs model was upgraded when aggregating institutional and community variables with HCR-20 or with RM2000V items, while other statistical models were not improved.

Liu and colleagues also built several predictive models using different statistical analyses including logistic regression, classification tree, and neural networks model using a sample of male prisoners in England and Wales released from 14th Nov 2002 to 7th Oct 2005 (Liu et al., 2011). A total of 1,363 released prisoners were interviewed and the degree of risk was measured by HCR-20 items. After dropping incomplete vases, the study sample is comprised of 1,225 male offenders who were released during the stud period.

The main outcome variable to be predicted by statistical models was violent re-conviction. 343 releasees (28.0 %) were re-convicted for violent crime, including homicide, major violence, minor violence, weapons offenses, aggravated burglary, and robbery. Although 383 cases were reconvicted for the non-violent crime after release, these individuals were classified into the non-violent reconviction group. 20 predicting variables were drawn from the HCR-20 risk assessment tool which is comprised of 10 historical factors, 5 clinical presentation items, and 5 future risk factors. To compare the predictive accuracy between the three models, the AUC was adopted instead of conventional parameters such as overall accuracy, specificity, or sensitivity. The AUCs of the NNs model using testing samples were slightly higher than LR or CART in all scenarios (.65-.70). However, when it comes to conventional parameters of performance, including sensitivity, specificity, and overall accuracy, there were no significant differences between the models. To summarize, it is concluded that the NNs model was moderately outperformed in predicting violent recidivism over logistic regression and classification and regression tree methods, although the difference is not significant.

Tollenaar and Van der Heijden (2013) built the predictive models of recidivism using samples in the Netherlands. As numerous machine learning algorithms were developed during the past few decades, the authors applied 11 machine learning-based predictive models to predict recidivism. The 11 types of machine learning based-model were as follows: Logistic regression, multivariate adaptive regression spline (MARS), linear discriminant analysis (LDA), flexible discriminant analysis (FDA), recursive partitioning (rpart), adaptive boosting (adaBoost), logitBoost, neural networks (NNs), linear support vector model (SVM), K-nearest neighbors (KNN), and partial least squares (PLS). The target population of the study was adult offenders who were found guilty in the Netherlands in 2005, thus 20,000 adult offenders were randomly selected out of 184,339 offenders in the total population. The performance parameters for the model evaluation used in the study were AUC, accuracy ACC, RMSE, SAR, overall calibration error, and local calibration error. The results of the analyses were generally consistent with the findings of previous studies. When predicting general recidivism, the predictive performance of the NNs model was moderately worse than the LR model, whilst NNs performs better than or similar to the other 9 models. When it comes to the violent recidivism prediction model, the NNs model performed significantly worse than LR and other predictive models except for rpart, logitBoost, and KNN. Unlikely, the NNs model outperformed LR in predicting sexual violence. However, some other models, such as LDA, PLS, and SVM outperformed NNs in almost all criteria.

One of the noteworthy findings in the study of Tollenaar and Van der Heijden (2013) was the result of clinical usefulness. The author argued that the ability of the model to predict the right class determines clinical usefulness. Since the accuracy varies by the choice of cut-off point and calibration of the probabilities, the authors arbitrarily set the sensitivity is equal to specificity. As Table 13 represents, the predictive accuracy of the NNs model was not the best among 11 machine learning models. While the NNs model showed significantly higher accuracy in predicting sexual recidivism than LR, the NNs model did not significantly better than LR in predicting general recidivism and violent recidivism. Overall, the NNs model performed moderately better than LR in predicting sexual recidivism, whereas there were no significant differences between NNs and LR when it comes to the prediction for general recidivism or violence recidivism.

Table 13

Accuracy of the Models When Sensitivity = Specificity, Test Data in Tollenaar and Van der Heijden (2013)'s Study

Model	Results for general recidivism	Results for violent recidivism	Results for sexual recidivism
Logreg	.704	.672	.587
MARS	.705	.676	.464
LDA	.705	.673	.660
FDA	.704	.676	.681
Rpart	.690	.653	.500
adaBoost	.696	.677	.523
logitBoost	.671	.645	.486
PLS	.705	.677	.705
KNN	.660	.624	.545
Nnet	.704	.662	.647
SVM, linear	.699	.671	.602

Note: Reprinted from Tollenaar and Van der Heijden (2013).

A similar comparative study between ML methods has been conducted by Hamilton et al. (2016). Using a relatively large sample of offenders in the state of Washington (N=297,600), the study examined the performances of recidivism prediction between ML models and conventional statistical analysis such as logistic regression, neural networks, random forests. In this study, each predictive model has been trained to predict four types of recidivism, such as felony, drug, violence, and sexual violence based on instruments of the Washington State Static Risk Assessment tool. The results of this study revealed that conventional statistical analysis, logistic regression, still yielded comparable performance to ML-based predictive models regardless of types of recidivism. In specific, when focusing on AUC, logistic regression is still the most efficient statistical analysis than RF and NNs in all types of recidivism.

Recently, similar attempts have been made to compare the predictive performance between ML-based models and conventional statistical analysis. Using 27,772 individual offender samples drawn from the data of Minnesota prison between 2003 and 2010, Duwe and Kim (2017) constructed 12 predictive models: simple logistic regression, full logistic regression, regularized logistic regression, Naïve Bayes, Decision tree, ANN, SVMs, Bootstrap Aggregating, Random Forests, LogitBoost, MultiBoosting, and LMTs. To compare the overall predictive performance between the 12 models, Duwe and Kim (2017) ranked each model according to the results of AUC, precision, recall rates, and ACC. According to their results of rank, LogitBoost recorded the highest rank with the highest precision (.544) and ACC (.823). On average, Random Forests and Multiboosting were ranked in the second-highest group.

While previous studies focused on the recidivism of released offenders, a study conducted by Ngo et al. (2015) applied ML methods to predicting inmate misconduct. The authors criticized the application of clinical risk assessment tools which has been established by generalized linear models for the sake of universal assessment without individual considerations. With this in mind, Ngo and his colleagues (2015) identified risk factors of inmate misconduct according to existing theoretical backgrounds, such as importation models, and test the accuracy of ML-based predictive models using those risk factors. The importation model contends that individual characteristics, such as gender, race, and socioeconomic status, which have been already determined before incarceration are associated with the maladjustment of inmates. Data for the study was drawn from the 2004 Survey of Inmates in State and Federal Correctional Facilities (SISFCF) conducted by the Bureau of the Census. After dropping incomplete cases, a total of 10,000 inmates incarcerated in State and Federal correctional facilities were randomly selected out of 10,328 valid cases.

Four types of statistical analysis were used to build the predictive models of inmate misconduct including logistic regression, CART, chi-squared automatic interaction detection (CHAID), and multi-layer perception NNs. The outcome variable, prison misconduct, was dichotomously measured by asking inmates with "Since your admission, have you been written up for or found guilty of breaking any prison rules?" To build the predictive model, 11 predicting indicators were identified, such as gender, age, race, marital status, a prior arrest, and age of the first arrest, according to the importation model. Ngo et al. (2015) calculated four common performance parameters (sensitivity, specificity, overall accuracy, and AUC) of each predictive model. When focusing on overall accuracy, there are no significant differences between the four models. However, the AUC of the NNs model is significantly higher than tree-based predictive ML methods (CART and CHAID), while no significant differences in AUC was found between LR and NNs. In line with the findings of previous studies, Ngo et al. (2015) concluded that the predictive performance of NNs is similar to that of conventional methods LR, while tree-based ML methods did not precisely predict antisocial behaviors than NNs and LR.

Summary for Application of Machine Learning to Predicting Antisocial Behavior.

Table 14

Summary for Application of Machine Learning to Predicting Antisocial Behavior

	Output	Method	Results
Gardner et al. (1996)	Violent behavior	Regression Tree	ML=NBR
Steadman et al. (2000)	Violent behavior	Classification Tree	ML ≥LR
Silver et al. (2000)	Recidivism	Classification Tree	ML<LR
Rosenfeld & Lewis (2005)	Violent behavior	Classification Tree	ML>LR
Neuily et al. (2011)	Recidivism	Random Forest	ML>LR
Pflueger et al. (2015)	Recidivism	Survival Random Forest	ML>LR
Caulkins et al. (1996)	Recidivism	Neural Networks	ML=LR
Palocsay et al. (2000)	Recidivism	Neural Networks	ML>LR
Yang et al. (2010)	Recidivism	Neural Networks, Tree	ML=LR
Liu et al. (2011)	Recidivism	Neural Networks, Tree	ML≥LR
Tollenaar & Van der Heijden (2013)	Recidivism	NNs, SVM, Tree	ML≥LR
Ngo et al. (2015)	Inmate Misconduct	NNs, Tree	ML=LR
Hamilton et al. (2016)	Recidivism	Neural Networks, RF	ML<LR
Duwe & Kim (2017)	Recidivism	Naïve Bayes, SVN, RF	ML>LR

In summary, it is still unclear that machine learning techniques can be used to forecast deviant behaviors and crime in the future. While ML techniques are widely accepted in many fields of science, it is rarely used in forecasting antisocial behaviors because of its low efficacy of prediction comparing to the conventional statistical methods. However, it is too early to conclude that the ML technique will not contribute to the studies of crime and deviant behaviors. Although the predictive performance of

ML methods is not constantly better than conventional statistical analysis such as Logistic Regression, ML methods can assist decision-makers to assess the risk of recidivism while considering relative costs of false-negative to false-positive (Berk, 2012).

CHAPTER III

Methods

Data

The data collected by the National Longitudinal Survey of Youth 1997 (NLSY97) will be used for this study. The NLSY97 is comprised of a nationally representative sample of 8,984 young men and women who were 12-16 years old when they first participated in the survey in 1997. The survey was conducted annually until 2011, and biennially from 2011 to 2017. The NLSY97 sample was selected to represent the population of the United States between the age of 12 and 16 years old in 1996 with 2,236 oversamples of Hispanics and non-Hispanic blacks. Youths and their parents both participated in the survey and answered questionnaires which cover nine general topics: Education, Training, and Achievement scores, Employment, Household, Geography, & Contextual Variables, Parents, Family Process & Childhood, Dating, Marriage & Cohabitation, Sexual Activity, Pregnancy & Fertility, Children, Income, Assets & Program Participation, Health, Attitudes, Expectations, Non-Cognitive Tests, Activities, and Crime & Substance Use.

The interviews were administered by an interviewer with the assistance of computer software. For the first wave of the interview in 1997, 96.8% of the interview was conducted in person, and only 3.2% of the interview was conducted via telephone. Until the fifth wave, most of the interview was administered in person (91.5% in 2001). However, the recent interview, which was held in 2017, was mostly administered via telephone (89.5%). Nearly 90 percent of the respondents were retained until the fourth round of the survey, however, 25 percent of the sample were dropped out of the survey in

2017. In addition, respondents who were 15 and 16 years old at the first survey are also not selected since the dependent variable, LCP offenders, are determined by analyzing the trajectory of criminal involvement between the age of 14 and 19. As a result, 5,419 individuals are used for the identification of LCP and AL offenders in the first step of the analysis. Missing values of cases were dealt with mode imputation because no systematic patterns of missing was detected.

Dependent Variable

Life Course Persistent Offender

The operationalization of life-course-persistent and adolescent offenders was varied by researchers. One common method for identifying LCP and AL offenders accommodates simple computerized algorithms (see Moffitt et al., 1993; 1996). Moffitt and colleagues (1993; 1996) developed a series of procedures to identify an individual's trajectory of offense history and divide them into LCP, and AL groups. First, they selected samples who showed high scores (i.e., one standard deviation above the mean score) of antisocial behavior scales more than three times within the four waves of the survey during childhood (i.e., at 5, 7, 9, and 11-year-old). Within the selected subsamples, they divided the samples into boys who had histories of antisocial behaviors (i.e., one standard deviation above the mean score) during adolescence (i.e., at 15 and 18-year-old) versus boys who did not. Individuals who met both criteria were designated as the LCP offenders. Individuals who had not been antisocial in childhood, but who showed high scores of antisocial behaviors scales at adolescence were assigned to AL offenders. Similarly, Nagin (1999) applied the algorithmic approach to identify three groups of individuals according to their histories of conviction; never convicted, AL, and chronic. In Nagin's (1999) study, individuals with the experience of conviction less than once during the study period, between 10 and 32, were assigned to the "never convicted" group. Next, the individuals who showed a history of conviction before their early 20s, but who stopped their offending after their early 20s are designated to AL offenders. The chronic group comprised individuals who had a history of conviction before their early 20s and continued engaging in criminal behaviors.

The second method for classifying individuals into LCP and AL offenders accommodates the longitudinal Mixture Modeling that is originally proposed by Jones, Nagin, and Roeder (1998). While the conventional growth trajectory analysis provides a single average growth trajectory of all individuals in the sample, Latent Growth Modeling (LGM)assumes that the growth trajectories of certain subsets of individuals are different from those of others in the sample (Jung & Wickrama, 2008). To separate the growth models for each group, the LGM approach estimates trajectories of groups according to given numbers of latent classes. In the previous study of Jones and his colleagues (2001), Latent Growth Model with four or five latent classes was most fitted in the sample of two longitudinal surveys in classifying trajectories of delinquent behaviors over time.

The NLSY 97 survey measures the level of criminal involvement by asking six questions administered by the children respondents each year from 1997 to 2011. Respondents were asked to answer how severely they were involved in the following criminal activities in the previous year: (1) Purposely damaged or destroyed property not belonging to the respondents, (2) Stole something worth less than $50, (3) Stole something worth $50 or more (including a car), (3) Other property crime, including fencing stolen property, possessing or receiving stolen property, or selling something for more than it was worth, (5) Attacked or assaulted someone, (6) Sold or helped to sell marijuana, hashish, or other hard drugs. Each item was answered with binary responses (0=No, 1=Yes). By summing scores of six items, the scale of criminal behavior of an individual at each wave is created. The scale of criminal behavior of individuals was generated per their age, therefore, a longitudinal series of behavioral scales during childhoods and young adults can be used to classify individuals to LCP and AL. Since the age of individuals at the entry of the survey varies from 12 to 16, not all individuals in the survey were included in this study. The individuals who were between 12 and 14 at the first wave were selected and their criminal behavior scales at the age of 14 through 19 were used for LCP and AL classification using the Latent Growth Modeling approach.

Independent Variables

Following the results of previous empirical studies on the influence of risk factors on persistent behavioral problems (see Moffitt et al., 1996; Moffitt & Caspi, 2001), three domains of risk factors, family, individual, and peer/environment, are included in this study. Family risk factor consists of a maternal relationship, family routines, parental monitoring, and maternal adolescent pregnancy. Individual risk factor comprises emotional problem, early onset of crime, belief in chances of getting arrested, experiences of going through hard times, physical/emotional strain, learning problems, use of substance, citizenship, race, and sex. Peer environmental risk factor consists of peer delinquency, exposure to gang members, housing environment, neighborhood environment, school environmental risk, and environmental safety. All risk factors except for the belief in chances of getting arrested and educational attainment are measured at the first wave of the survey. The degrees of belief in chances of getting arrested and use of substance are measured when the respondents are at the age of 14.

Table 15

Lists of Risk Factors

	Name of Factors
Familial Risk Factors	Maternal Relationship
	Family Routines
	Parental Monitoring
	Maternal Adolescent Pregnancy
Individual Risk Factors	Emotional Problem Index
	Early Onset of Crime
	Belief in Chances of Getting Arrested
	Hard Times
	Physical/Emotional Strain
	Learning/Emotional Problem
	Use of Substance
	Citizenship
	Race
	Sex
Peer/Environmental Risk Factors	Delinquent Peers
	Exposure to Gang Members
	Housing Environment
	Neighborhood Environment
	School Environment
	Environmental Safety

Familial Risk Factors

Maternal Relationships. The level of relationships with their mother or caregivers was measured by 8 items. Respondents were asked to answer the following questions: (1) " I think highly of her", (2) "She/he is a person I want to be like", (3) "I really enjoy spending time with her", (4) How often does she praise you for doing well", (5) "How often does she criticize you or your ideas?", (6) "How often does she/he help you do things that are important to you?" (7) "How often does she blame you for her problems?", (8) "How often does she make plans you and cancel with respondent?' Respondents were asked to respond on 5-point scale (0 = "strongly disagree", 1 = "disagree", 2 = "neutral or mixed", 3 = "agree", 4 = "strongly agree"). Eight items were summed together and normalized to create a measure of *maternal relationships*. Higher scores represent the closer relationship with the respondent's mother or caregivers.

Family Routines. Levels of routine activities with family were measured by four items administered by children respondents. Respondents were asked to answer how many days from 0 to 7 they involve in the following activities in a typical week: (1) Eating dinner with your family, (2) Doing housework gets done when it is supposed to, for example cleaning up after dinner, doing dishes, or taking you're the trash, (3) Doing something fun as a family such as play a game, go to a sporting event, go swimming and so forth, (4) Doing something religious as a family such as go to church, pray or read the scriptures together. The answers were responded to with an 8-point scale. All answers to items were summed together to create a measure of family routines. Higher scores indicate more days spent in routine activities with the family.

Parental Monitoring. The level of *maternal/paternal monitoring* was measured by eight items administered by respondents. Respondents were asked to answer four questions regarding what degree do they feel their residential mother and residential father monitor them respectively: (1) How much does he/she know about your close friends, that is, who they are?, (2) How much does he/she know about your close friend's parents, that is, who they are?, (3) How much does he/she know about who you are with when you are not at home?, (4) How much does he/she know about who your teachers are and what you are doing in school?. Each item was answered by an ordinal response ranging from 0 to 4. The measure of maternal monitoring (0 to 16) and fraternal monitoring (0 to 16) were generated by totaling the scores, then, the *parental monitoring* measure was created by summing two measures together. Higher scores represent higher levels of parental monitoring.

Maternal Adolescent Pregnancy. The mother's age at the birth of the child was administered by the parents of respondents. If the biological mother of the respondent were under the age of 18 when the respondents were given birth, the respondent was coded to 1 (=Adolescent Pregnancy).

Individual Risk Factors

Emotional Problem Index. The level of *emotional problems* was measured by four items administered by children respondents. Respondents were asked to answer the following questions: (1) "Your school work is poor", (2) "You have trouble sleeping", (3) "You lie or cheat", (4) "You are unhappy, sad, or depressed". Answers to each question were responded to on three-point scales ranging from 0 to 2. The *emotional problem index* was generated by totaling all scores of four items.

Early Onset of Criminal Behaviors. Every year, respondents are asked to mark their monthly arrest status in the previous year. Based on the self-reported experience of arrest during the survey, the first arrest date of respondents was estimated by interviewers. The measure of *early onset of criminal behavior* is recorded on the dichotomous response (0=No early-onset, 1=Early onset). Respondents who have been arrested before they reach 13 years old are designated to the early onset group (=1). Regardless of the experience of arrest after their age of 13 years old, other respondents who have not been arrested before the age of 13 are designated to the "No early-onset" group (=0).

Belief in Chances of Getting Arrested. Individuals may have different levels of expectation of getting caught when doing criminal behaviors. According to the rational choice theory, the low expectation of getting caught by law enforcement may increase the chance of committing criminal behaviors when the benefit from the crime exceeds the risk. Respondents were asked to answer at the age of 14: "What is the percentage chance that you would be arrested if you stole a car?". Answers to the question were responded to on 11-point scales ranging from 0 (0 percent) to 10 (100 percent).

Individual Strain. Individual experiences of stressful events were measured by three individual items. First, whether respondents have gone through *hard times* during their lifetime before the first wave of the survey were measured by asking: "Have you ever lived through hard times?". The measure of *hard times* was responded on a dichotomous scale (0 = No, 1 = Yes). Second, respondents' *physical/emotional strain* was measured by asking their parents or caregivers: "Does your child have or has he/she ever had any physical, emotional, or mental condition that limits or has limited ability to attend school regularly, do regular schoolwork, or work at a job for pay?”. The measure of *physical/emotional strain* was responded to on a dichotomous scale (0 = No, 1 = Yes). Third, respondents’ *learning/emotional problem* was also measured by asking their parents or caregivers: “Does your child now have or has he/she ever had a learning or emotional problem that limits or has limited the kind of schoolwork or other daily activities he/she can perform, the amount of time he/she can spend on these activities or his/her performance in these activities?” The measure of *learning/emotional problem* was responded to on a dichotomous scale (0 = No, 1 = Yes).

Use of Substance. The degree of using the substance at the age of 14 was measured by asking three items: “Have you ever smoked a cigarette”, “Have you ever had a drink of an alcoholic beverage?”, and “Have you ever used marijuana?” Three items were responded to on dichotomous answers. A single measure, *use of substance*, by aggregating three items (range 0 to 3).

Citizenship. The citizenship status of respondents was administered by the parents of respondents. Respondents who were not born in the U.S. territory or did not have citizenship at birth are coded to 0 (=Non-citizen).

Race. Three dummy variables regarding the race of respondents were generated to (1) Black (1=White, 0=Non-Black), (2) Hispanic (1=Hispanic, 0=Non-Hispanic), and (3) other race (1=other, 0=Hispanic or Black)

Sex. Sex was coded such that male =1 and female =0.

Peer and Environmental Risk Factors

Delinquent Peers. The degree of delinquency of peers was measured by 6 items administered by children respondents. Respondents were asked to estimate the percentage of peers in their grade involved in the following types of delinquent behaviors : (1) Belong to a gang that participates in illegal activities, (2) Cut classes or skip school, (3) Ever use marijuana, inhalants, or other drugs, (4) Get drunk at least once per month, (5) Smoke cigarettes, (6) Ever had sexual intercourse (asked of those age 15 and older). Respondents were asked to respond on 5-point scale (1 = Almost none, 2 = About 25%, 3 = About half, 4 = about 75%, 5 = Almost all). All six items were summed together and normalized to create a measure of delinquent peers. Higher scores represent higher levels of peers' delinquency.

Exposure to Gang Members. The degree of exposure to gang members was measured by two items administered by children respondents. Respondents were asked to answer: "Are there any gangs in your neighborhood or where you go to school?", and "Do any of your brothers, sisters, cousins or friends belong to a gang?" Respondents were asked to respond on dichotomous answers for each item (0 = No, 1 = Yes). A single measure, *exposure to gang members*, was created by aggregating two items.

Physical Environmental Risk. Two dimensions of physical environmental risk were independently measured. Levels of the *housing environment* were measured by two items. The items were measured by field observation of the interviewer while the interview was administered in the house of respondents. Interviewers recorded the status of the physical environment of the house according to the following questions: (1) "How well kept in the interior of the home in which the youth respondent lives?", and (2) "How well kept is the exterior of the housing unit where the youth respondent lives?" Both items were responded to on a three-point scale (1 = very well kept, 2 = fairly well kept but some evidence of needed repairs, 3 = poorly kept). The *housing environmental risk* index was created by tallying scores of two items (range 2 – 6). Higher scores indicate a riskier housing environment.

The degree of the *neighborhood environment* was measured by one item. Interviewers were asked to answer the following question: “How well kept are most of the buildings on the street where the adult/youth resident lives?” The item was responded to on a three-point scale (1 = very well kept, 2 = fairly well kept but some evidence of needed repairs, 3 = poorly kept).

School Environment. Levels of school environment were measured by five items administered by children respondents. Respondents were asked to think about their school in general and answered how much do they agree with each of the following statements about their school and teachers: (1) The teachers are good, (2) The teachers are interested in the students, (3) Students are graded fairly, (4) Discipline is fair, (5) I feel safe at this school. Respondents were asked to respond on four-point scale (1 = Strongly agree, 2 = Agree, 3 = Disagree, 4 = Strongly disagree). Answers are summed together to generate the measure of the school environment.

Environmental Safety. The degree of the neighborhood and home safety was measured by one item administered by the interviewer. Interviewers recorded their feelings when they went to the respondent’s home for the interview in the survey year of 1998. Interviewers were asked to answer: “When you went to the respondent’s neighborhood/home, did you feel concerned for your safety?” The item was responded to on a dichotomous answer (0 = No, 1 = Yes).

Analytic Strategy

The cohort of the NLSY97 study consists of 8,984 American youth born between 1980 and 1984. Consequently, the ages of respondents vary at each wave. Since the first step of the present study is to distinguish LCP and AL offenders according to individual trajectories of offenses by age, all variables should be newly created according to the age of the respondents. For instance, the criminal involvement scale in 1997 for the respondents who were born in 1984 should be renamed as “criminal involvement at the age of 13”, while the same scale in 1997 for the respondents who were born in 1982 should be renamed as “criminal involvement at the age of 15”. Using Stata 15.0 software, all study variables were renamed with the age of the respondents at each wave.

Before the analysis, all cases with the missing values of the criminal involvement scale at least one time were dropped out of the sample to avoid the null effect on prediction accuracy. After dropping the cases with missing values of the criminal involvement scales, 5,419 cases were left in the sample. When it comes to the cases with missing values of risk factors, the value of mode of each factor was imputed.

In the first step of the analysis, each individual’s life-course trajectory of criminal involvement is determined by the Latent Growth Modeling approach as presented by Jones and his colleagues (2001) using package *flexmix* on R 3.5.2. Each respondent is classified into one of the groups, LCP, AL, and low offense group, according to their patterns of the trajectory of self-reported criminal involvement scales. After the clustering process of LCP and AL is done, the binary outcome factor, Life-Course Persistent offender, is generated. Based upon the results of Latent Growth Modeling, two groups, LCP (=1) and AL (=0) offenders, are generated and included in the forecasting analysis.

The overfitting issue may occur if the trained models predict the outcome using the same dataset as the model was being trained. Therefore, the performance of predictive models built up with the training data should be examined using discrete datasets that were not used during the training process. To avoid the overfitting issue, the entire data should be divided into the training and the testing datasets before the training process. Thus, before the analysis, 60 percent of the sample are randomly assigned to the training data and the rest of the sample is assigned to the testing data.

In the second step of the analysis, using seven different types of machine learning-based forecasting models, the predicting models of LCP will be constructed using all study variables: Logistic Regression, Decision Tree, Recursive Partitioning, Naïve Bayes, Random Forest, single-layered Neural Networks, and Deep Learning. After the training process using training data is over, the predictive performances of six ML-based predictive models are compared to those of the conventional logistic regression model using testing data. All forecasting models will be built upon the corresponding packages in R 3.5.2. All independent variables, risk factors in this study, are standardized for the better performance of the training process.

Logistic Regression

Logistic Regression (LR) is one of the generalized linear models that uses a logistic function, also called the sigmoid function, to model a binary dependent variable. LR is also used for the classification by estimating the log odds of each case. By transforming the log odds to the probability score, the outcome is classified as a positive class when the probability is close to 1 and as a negative class when the probability is close to 0. In this study, the simple LR function is used to estimate the probability, and the cutoff value of .5 for the classification is adopted. For the analysis, the *"glm"* function in R package was used.

Classification and Regression Tree

All tree-based classification models follow the approach of a decision tree algorithm which is a tree-like flowchart structure that is comprised of multiple nodes and branches. Each node of the decision tree has two burst nodes (splitting paths) and one path for each observation is selected by the decision rules of each node. The goal of tree-based classification models is to classify observations into a subset of the sample with homogeneous characteristics (Berk & Bleich, 2014). In decision tree-based classification methods, as proceeding to the next terminal nodes, the variation of the outcome within each child node becomes smaller. The tree-based classification model varies by type of separation criteria, stopping rules, and pruning methods. Each subset of the data partitioned by previous criteria is called a node, and the nodes at the bottom are called terminal nodes. To construct decision trees, the following algorithm is applied.

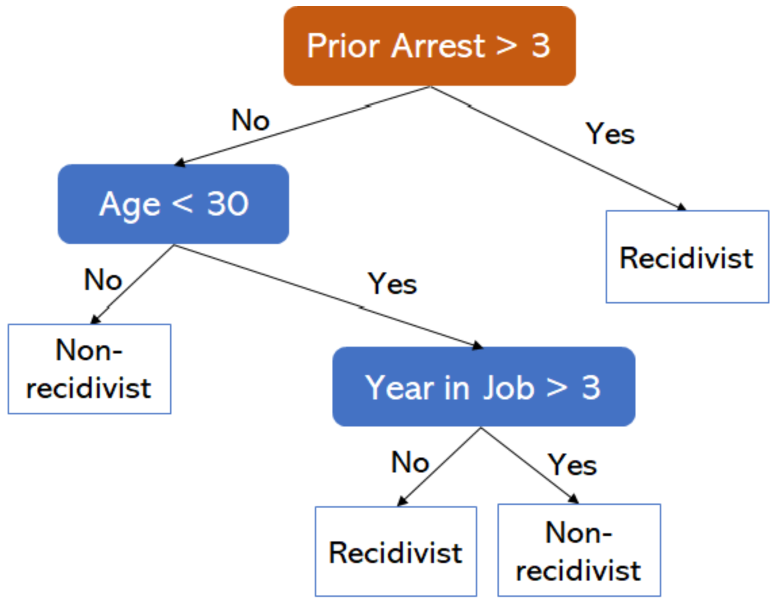
1. Start at the root node.

2. For each predictor variable, find the set of node that minimizes the sum of the node impurities in the two child nodes and choose the split that gives the minimum overall predicting variable and nodes.

3. If a stopping criterion is reached, exit, otherwise, apply step 2 to each child node in turn. (Loh, 2011, p. 14)

Figure 7

Example of Tree-Based Classification Methods



According to the type of stopping criterion and ways of calculating node impurities, several different tree algorithms have been developed. A *Classification and Regression Tree (CART)* model is a branch of tree-based classification model that uses *Recursive Partitioning* (Breiman et al., 1984). CART includes two types of decision trees, classification tree, and regression tree. A classification tree is designed to predict the discrete (categorical) outcome, whereas the regression tree is designed to predict the continuous outcome. According to the pruning process and splitting criteria, several advanced decision tree models can be made. CART is distinguished from other tree methods by using the greedy approach because at each step of splitting partitioning the best outcome is selected then no revision is made later (Berk, 2012).

CART uses a generalization of the binomial variance, also called a Gini index, for its node impurity function. Gini index is calculated by summing all squared probabilities of each class, which indicates how much every specification directly influence subsequent case.

$$\text{Gini} = 1 - \sum_{i=1}^{c} (p_i)^2$$

Where pi is the probability of an object being classified to a particular class (Breiman et al., 1984).

To construct CART, the following algorithm is applied.

1. The single variable is found which best splits the data into two groups.

2. Another single variable is found which best splits the data into two groups for each subgroup.

3. Recursive steps split the data into two subgroups until the subgroups either reach a minimum size or until no improvement can be made.

4. A cross validation process trims the full tree to build subsets of trimmed tree.

5. Estimate the risk for each trimmed tree and the best trimmed tree with the lowest risk is selected. (Therneau, Atkinson, & Foundation, 2011, p .12).

There are two commonly used packages in R to run the CART model, “rpart” and “tree”. Fundamentally, two packages are developed on the Breiman and colleague’s (1984) book, but there are differences in the way of treating missing values and providing tuning parameters. Both packages, “rpart” and “tree”, will be used in this study to compare the results.

Naïve Bayes

Naïve Bayes classifier, also called the Bayesian classifier, is a conditional probability model based on the Bayes theorem. Although it is unrealistic, the classifiers are simplified by assuming that each factor is independent of the value of other factors in a given class. Despite the simple algorithm and unrealistic assumption, the classifier is known to outperform other sophisticated classification algorithms (Rish, 2001).

Formula for Bayes' theorem

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$$

Where:

$$P(A) = The\ probability\ of\ A\ occurring$$

$$P(B) = The\ probability\ of\ B\ occurring$$

$$P(A|B) = The\ probability\ of\ A\ given\ B$$

$$P(B|A) = The\ probability\ of\ B\ given\ A$$

Based on the Bayes theorem, the classifier calculates the posterior probability of class given predictors.

$$P(c|x) = \frac{P(c) \cdot P(x|c)}{P(x)}$$

$$P(c|X) = P(x_1|c) * \ P(x_2|c) * P(x_3|c) * \dots * \ P(x_n|c) * P(c)$$

Where:

$$P(c) = The\ prior\ probability\ of\ class\ (c)$$

$$P(x) = The\ prior\ probability\ of\ predictor\ (x)$$

$$P(x|c) = The\ probability\ of\ predictor\ x\ given\ class$$

$$P(c|x) = The\ posterior\ probability\ of\ class\ given\ predictor\ x$$

$$P(c|X) = The\ final\ posterior\ probability\ of\ class\ for\ each\ case\ considering\ all\ predictors$$

For each predictor, the posterior probability of all classes (e. g. fail or no fail) are calculated according to the Bayes theorem. Next, the final posterior probability of class for each case is calculated and the case is classified into the class which scored higher posterior probability. For instance, if the standardized final posterior probability of fail for the case ($P(Fail|X)$) is .2 and the standardized final posterior probability of no-fail for the case ($P(No\ fail|X)$) is .8, the case is classified as no-fail. The *naive Bayes* function of package *"e1071"* in R 3.5.2 was used to build the Naïve Bayes classifier-based predictive model.

Random Forest

Random Forest is one of the ensemble models that generate several classifiers and aggregate the results to make the final decision. Ensemble models in ML refer to the classifying algorithms combining several machine learning techniques into the final predictive model (Breiman, 2001). In case of RF, to build several trees and select classifiers for each tree, subsamples are randomly selected and a subset of predictors is also randomly chosen (Breiman, 2001). Comparing to other classification methods, the random forests model is well known for its performance and overfitting issues (Breiman, 2001; Berk, 2012). The random forests algorithm is as follows.

1. Select ntree bootstrap samples from the data

2. For each samples, grow an unpruned classification or regression tree with the following modification: at each node, randomly sample mtry of the predictors and choose the best split from among those variables.

3. Predict new data by aggregating the predictions of the ntree trees based on majority votes for classification, average for regression. (Liaw & Weiner, 2002, p. 18)

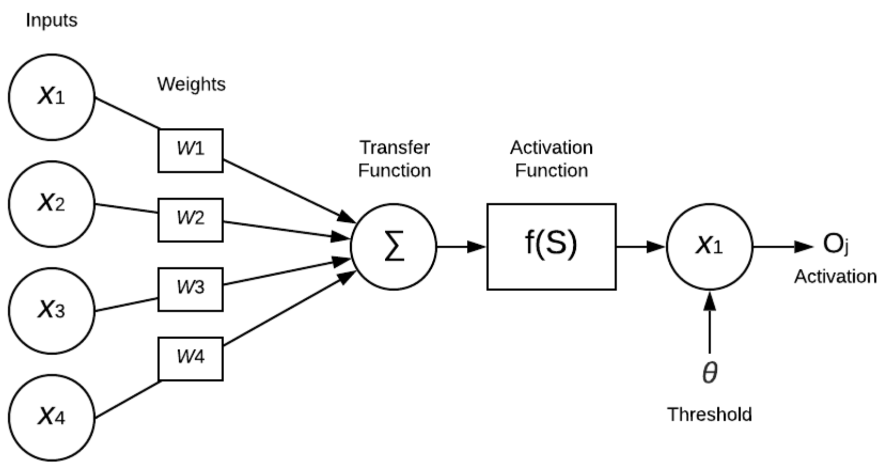
For the analysis, randomForest function of *"randomForest"* package in R 3.5.2 is used to train the random forest model.

Neural Networks

Neural Networks are computational algorithms imitating the human nervous system in making decisions. As the human nervous system receives the stimulations and processes its decision via complex networks of neurons, the NNs receives the information via the input layers and generates artificial neurons (perceptrons) to predict the output. The layers of generated nodes, also called hidden layers, are connected to the nodes of the previous layer and the nodes of the subsequent layer. Figure 8 represents how the perceptrons are connected to other nodes. Each node in the hidden layers has an activation function that defines whether the node is activated or not, based on the value of summarized net input. Once the output value of the activation function reaches a certain threshold, the nodes are activated.

Figure 8

Algorithms of Neural Networks Analysis



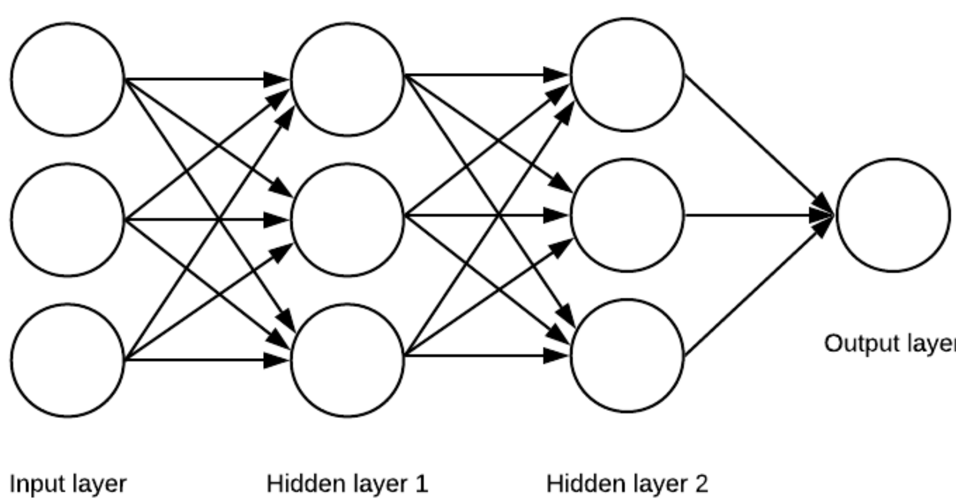
NNs model can be comprised of a single layer of perceptrons and multiple layers of perceptrons. NNs with multiple layers of perceptrons are called *Deep Learning*. Minsky and Papert (1969) introduced a single-layer neural networks model. In the single-layer NNs, one layer of nodes is generated to predict the optimal output for the given input values. Rumelhart et al. (1986) advanced the NNs with the back-propagation algorithms that can train the perceptrons of the hidden layer based on the output layer's error. NNs model has several advantages over multiple linear regression. First, the NNs does not require the assumption of the normal distribution of errors (Caulkins et al., 1996). Second, the NNs model utilizes the iterative process (Warner & Misra, 1996). Third, the NNs model has a high tolerance for noise and is suitable for nonlinear relationships (Razi & Athappilly, 2005). Despite the advantage of the NNs model, however, it is very sensitive to the tuning parameters and has overfitting issues (Hastie et al., 2009). The *nnet* function of library "nnet" in R 3.5.2 was used to build the single-layered NNs model.

Deep Learning

Deep learning refers to the NNs model with 2 or more hidden layers of perceptrons as presented in Figure 9. When building *Deep Learning* models, the number of hidden layers and hidden layers in each layer affects the efficiency and accuracy of the classification results. The fine-tuning process, therefore, is inevitable but challenging step to optimize the deep learning model. To provide the best fitted and accurate model, an iterative method with sequential numbers of layers and perceptrons in each layer is applied to find the best-fitted model for the data. The *"h2o"* package in R was used for building a *Deep Learning*, NNs with multiple layers of perceptrons, model.

Figure 9

Example of Multilayer Neural Networks model



CHAPTER IV

Results

Identification of Life-Course-Persistent Offenders

To identify the individuals showing high levels of criminal involvement throughout their adolescence, the Latent Class Growth Analysis was adopted. LCGA is a type of Latent Growth Modeling that identifies different trajectories of criminal behaviors of adolescents over years as used in the previous studies (Jones et al.,2001). Using *flexmix* package in R 3.6.1 software, five models with a different number of latent classes were estimated. Table 16 shows the results of each model with given numbers of latent classes. The Bayesian information criterion (BIC) is used to select the optimal model in line with the previous studies (see Nagin, 1999). The closer the value of BIC of the model represents the better model (Keribin, 2000). According to the results of the BIC of five models with a different number of latent classes, the model with three latent classes is the most optimal (BIC = -72848), while the model with one latent class is the worst fitted model (BIC= -83607). In addition, the models with four and five latent classes identify one or more classes with no observation as can be seen in Table 17, hence the model with three latent classes is used for identifying LCP and AL offenders.

Table 17 displays the outcome of the clustering process according to LCGA. A group called “low offense” is composed of individuals who barely have committed criminal behaviors between the age of 14 and 19. About 84 percent (n=4,555) of the sample are classified as the “low offense” group. The second group called “Adolescent Limited” is composed of individuals who show relatively high levels of criminal involvement during their adolescence but desist from crime as they are getting older. This group is estimated to make up about 12 percent (n=676) of the sample population. The third group called "Life-Course-Persistent" consists of individuals who show the highest levels of criminal behaviors during their adolescence and constantly engage in criminal offenses. About 3 percent (n=188) of the sample are classified into this LCP group. Overall, it is shown that the proportion of each group is analogous to the results of previous studies (see Moffitt, 1993; Nagin, 1999).

Developmental trajectories of criminal involvement for each group are shown in Figure 10. As expected, the mean value of the criminal involvement scale of the LCP group is constantly higher than the other two groups. The average criminal involvement scale of the AL group also shows respectively high levels of at the of 14, however, they stopped showing problematic behaviors before the age of 18 in general.

Table 16

Model Fit Indices for Five Models

Model	Number of Class	Log Likelihood	AIC	BIC
1	1	-41788	-83607	-83607
2	2	-37161	-74384	-74530
3	3	-36205	-72503	-72848
4	4	-36205	-72534	-78655
5	5	-336205	-72566	-81477

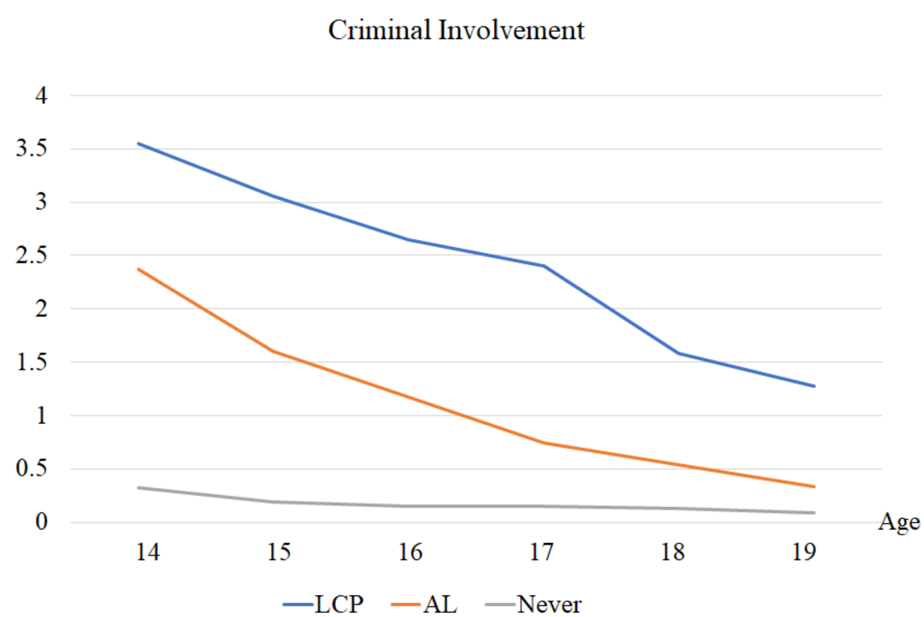
Table 17

Number of Cases in Each Class

	Model 3	Model 4	Model 5
Class 1	676	775	775
Class 2	4,555	4,456	4,456
Class 3	187	187	187
Class 4	-	0	0
Class 5	-	-	0

Figure 10

Observed Mean of Criminal Involvement for Each Group per Age



Forecasting Life-Course-Persistent Offenders

To meet the goal of the current study, individuals categorized as a low offense group are not used to build the forecasting model. To be specific, 187 LCP offenders and 676 AL offender samples are utilized to train and test the forecasting model of LCP using

ML and logistic regression analysis. A binary outcome variable indicating whether the individual is a life-course-persistent or adolescent limited offender is generated. As mentioned above, 60% of the total sample are randomly selected for the training process and the remaining samples are only used to validate the trained model to avoid overfitting issues. Therefore, a subsample with sample size 518, including112 LCP and 406 AL offenders, are used to train the forecasting model, while the other subsample with sample size 345 comprising 75 LCP, and 270 AL offenders are used to evaluate the trained ML-based forecasting model.

The 21 risk factors of three domains are used to train the forecasting models. Table 18 presents the results of descriptive statistics of all risk factors for the entire sample (N = 5,419). To validate the forecasting efficiency, the Area Under Curve (AUC) of each model is compared. Further, the overall accuracy, specificity, and sensitivity of each model are also compared to evaluate the performance of the prediction.

Table 18

Descriptive Statistics of Risk Factors

Name	**Total**	**Group Mean**		***p***
	Mean	**LCP**	**AL**	
Family Factor				
Maternal Relationship	24.12	21.40	22.05	.29
Family Routines	14.91	13.06	13.52	.32
Parental Monitoring	15.84	12.21	13.30	.05
Maternal Adolescent Pregnancy	.13	.09	.16	.04
Individual Factor				
Emotional Problem	2.15	3.11	2.81	.03

(continued)

Early Onset	.02	.06	.04	.27
Belief in Chance of Getting Arrested	6.00	1.62	1.67	.58
Hard Times	.05	.03	.07	.06
Physical/Emotional Strain	.07	.09	.09	.95
Learning/Emotional Problem	.11	.17	.15	.56
Use of Substance	.75	1.75	1.51	.01
Citizenship	.76	.77	.78	.59
Black	.26	.20	.22	.50
Hispanic	.21	.23	.19	.19
Sex (male)	.52	.79	.66	.00
Peer and Environmental				
Delinquent Peers	9.63	12.53	11.63	.05
Exposure to Gang	.29	.65	.51	.00
Housing Environment	3.02	3.11	3.14	.83
Neighborhood Environment	1.58	1.53	1.62	.09
School Environment	9.64	10.67	10.33	.12
Environmental Safety	.11	.07	.13	.03

Note. Total N = 5,419, AL group n=676, LCP group n=188

The baseline model for the reference to examine the predictive performance of the ML-based forecasting model is the Logistic Regression-based forecasting model. Similar to the conventional explanatory modeling approach, coefficients and the intercept of the equation are determined using training data. However, only the testing data is used for the validation process. Once the logistic regression equation is determined, the probability of being classified as LCP of each case in the testing dataset can be calculated. If the calculated probability of being LCP is equal to or higher than .50, the case is classified as LCP, and vice versa. The default cutoff value of log odds, 0 (probability = .50), is used because the results of tuned models using a different cutoff value of the log odds showed no difference in performance parameters.

Table 19 shows the confusion matrix of forecasting results of LR based model. The overall accuracy of the LR-based forecasting model is about 80%, in other words, the LR model misclassified samples into the wrong group by 20% of chance. However, when it comes to the true positive rate, only 27% (20 out of 75) of actual LCP offenders were correctly forecasted as persistent offenders. Considering that the goal of the forecasting model is to identify the high-risk offenders at earlier stages, missing a large portion of high-risk individuals in the future as the LR-based model does would not be welcomed by juvenile justice administrators or practitioners.

Table 19

Confusion Matrix of Logistic Regression Model

	Forecast AL	Forecast LCP	Model error
Observed AL	257	13	.05
Observed LCP	55	20	.73
Use error	.18	.39	Overall = .80

Two CART models were built by *tree* and *rpart* package. CART model is a relatively conventional model than other ML techniques in the current study. The overall accuracy of the CART (*tree*) model is reported as 82%, in other words, only 18 percent of the testing sample are classified into wrong groups. However, similar to the issue of the LR-based forecasting model, the model fails to forecast about 69 percent of chronic offenders (52 out of 75). The high overall accuracy of the model is derived from classifying individuals into the majority group, Adolescent Limited offenders, as many as possible.

Table 20

Confusion Matrix of CART (Tree) Model

	Forecast AL	Forecast LCP	Model error
Observed AL	260	10	.04
Observed LCP	52	23	.69
Use error	.17	.30	Overall = .82

The second CART model was built using the *rpart* package. Table 21 presents the results of classification using the CART (*rpart*) model. As can be seen in Table 21. the overall accuracy of the CART (*rpart*) model is 78%, while only 15 percent of the LCP offenders are correctly forecasted by the model. The overall accuracy and true negative rate are relatively high because a majority of samples are classified into the AL group.

Table 21

Confusion Matrix of CART (rpart) Model

	Forecast AL	Forecast LCP	Model error
Observed AL	259	11	.04
Observed LCP	64	11	.85
Use error	.20	.50	Overall = .78

The third ML-based model was built based on the Naïve Bayes modeling approach. As can be seen in Table 7, the overall accuracy of the forecasting model is about 67% which means that about 33% of the testing sample are misclassified into the wrong groups. However, unlike LR or RPART model, the true positive rate of the model significantly increased by sacrificing the overall accuracy. To be specific, the forecast of chronic offenders succeeds by 60% of chance (45 out of 75).

Table 22

Confusion Matrix of Naive Bayes model

	Forecast AL	Forecast LCP	Model error
Observed AL	187	83	.31
Observed LCP	30	45	.40
Use error	.13	.65	Overall = .67

The fourth ML-based model was built by using Random Forests modeling. Unlike the previously examined ML-based models, there is room for the modification of the training process in the RF model to yield the best performance as intended by the researcher. In this study, the balanced RF approach is adopted as suggested by Berk (2012). The balanced model refers to the modeling approach selecting more cases of the minority group, LCP group in this study, during the training process. Since the ensemble model such as RF repeatedly subsamples its training sample out of the training sample and determines whether each case fails (=0) or no-fails (=1) by aggregating results of all base models, researchers may decide the number of fail and no-fail cases for each sub-training model. To increase the True Positive Rate, 20 fail samples (AL) and 20 no-fail samples (LCP) were selected to train each base model in this study. Table 23 represents the results of the balanced RF model. Despite the low overall accuracy of the model, the TPR of the RF model is significantly higher than that of previous tree-based models, two CART models. To be specific, the LCP offenders are successfully forecasted by 59% (44 out of 75) chance. On the other hand, the overall accuracy of the model is 67 percent.

Table 23

Confusion Matrix of Random Forests Model

	Forecast AL	Forecast LCP	Model error
Observed AL	188	82	.30
Observed LCP	31	44	.41
Use error	.14	.65	Overall = .67

The fifth ML-based model was built based on single-layered Neural Networks (NNs) model with 1,000 maximum number of iteration. During the fine-tuning process of the NNs model, the number of perceptrons and the maximum number of iteration may be modified by researchers. In this study, 99 different single-layered NNs models were built according to the number of perceptrons, from 2 to 100, and the results of each model were compared to select the best-performed model. Among 99 models, the model providing the highest AUC for the validating sample is selected. As consequence, the model with three perceptrons is selected according to the performance parameters. Table 24 presents the results of forecasting using the NNs model. The overall accuracy of the model is reported as 65%, while the TRP is about 39%. Although it is common to apply the K-fold cross-validation to select the best-performed model, the validation process is not applied because the reliability of the validation is significantly low when the size of the training dataset is small.

Table 24

Confusion Matrix of Neural Networks Model

	Forecast AL	Forecast LCP	Model error
Observed AL	177	93	.34
Observed LCP	29	46	.39
Use error	.14	.67	Overall = .65

The last ML-based model was built based on the Deep Learning technique. Deep Learning refers to the Neural Networks model containing more than one layer of latent perceptrons. Among several advanced Deep Learning packages in R, the *h2o* package is applied to build the Deep Learning model. Similar to the single-layered NNs model, the number of perceptrons may be selected arbitrarily by researchers. Further, the number of layers of perceptrons and times epoch, the number of passes of the training process, are also selected by researchers. Upon the results of a fine-tuning process, the model with two hidden layers, 12 perceptrons for the first layer and 6 perceptrons for the second layer, and 12 times epoch is selected. As can be seen in Table 25, the overall accuracy of the final model is reported as 69%, while the true positive rate is about 75% (49 out of 75).

Table 25

Confusion Matrix of Deep Learning Model

	Forecast AL	Forecast LCP	Model error
Observed AL	189	81	.30
Observed LCP	26	49	.35
Use error	.12	.62	Overall = .69

Performance Evaluation

To determine which predictive model performs best, several parameters such as AUC, sensitivity, precision, specificity, and F1 score should be examined as well as the test accuracy and TPR mentioned above. To compare the overall predictive performance of the forecasting models, Table 26 presents a summary table including all performance parameters of LR and ML-based forecasting models. Among several performance measures commonly used for evaluating the predictive performance of the model, AUC is well known for judging the overall predictive power of the tool (Bradly, 1997). In this study, the AUC of the Random Forests model is reported the highest among the seven models. Specifically, the Random Forests model is the most efficient for distinguishing LCP and AL offenders among the testing samples. Considering that the model with higher than .70 of AUC can be considered acceptable (Hosmer & Lemeshow, 2000), three models, Logistic Regression, Random Forest, and Deep Learning, are only acceptable to use to predict LCP offenders.

Other overall predictive performance measures are presented in Table 26. As can be seen in Table 26, CART (*rpart*) provides the highest specificity, and CART (*tree*) provides the highest Accuracy, Precision, and Specificity among all models. However, the test accuracy of the model is not always a good parameter when using imbalanced data as the data used in this study is, since the null accuracy of the testing data is 78% (270 out of 345) if the model intentionally labels all cases as AL offenders. Especially for the stakeholders of the juvenile justice system, missing actual high-risk individuals due to misclassification may be more costly than labeling actual low-risk individuals as an LCP group. Moreover, relatively high scores of Precision and Specificity of the two models are derived from sacrificing the sensitivity score, therefore, it is hard to tell that two CART models perform better than other models in predicting LCP offenders. Meanwhile, Deep Learning provides the highest Sensitivity and F1-score, and Random Forests provide the highest AUC.

Considering the different costs between false negative and false positive from the stakeholders' perspective, Sensitivity also called Precision can be a good evaluation metric. According to Table 26, the Deep Learning model delivers the highest Sensitivity scores (=.653), while the CART (*rpart*) model records the lowest (.146). If the focus of the predictive model is identifying as many high-risk juveniles as possible, Deep Learning can be the best predictive model.

As can be seen in a summary table, the predictive performance of all predictive models is comparable to some extent. There seem no perfect ML-based predictive models that significantly outperform LR. Meanwhile, no performance measure of Logistic Regression is reported the highest among seven predictive models. To be specific, the sensitivity of the LR model is significantly lower than those of all ML models except for the CART (*rpart*) model. All in all, since the goal of this study is to find the best predictive model for identifying the LCP offenders, Random Forest and Deep Learning performed better than other models.

Table 26

Summary Table for Predictive Performance Metrics

	Logistic Regression	Decision Tree	RPART	Naïve Bayes	Random Forests	Neural Networks	Deep Learning
Accuracy	.803	**.820**	.783	.672	.672	.646	.690
AUC	.708	.602	.672	.657	**.717**	.666	.707
Sensitivity	.267	.307	.146	.600	.587	.613	**.653**
Precision	.606	**.697**	.500	.351	.349	.331	.377
Specificity	.952	**.963**	.959	.693	.696	.656	.700
F1-score	.370	.426	.227	.443	.438	.430	**.478**

Evaluating the Importance of Risk Factors

Upon the evaluation of predictive performance between models, it is revealed that the Random Forests and Deep Learning models slightly outperform Logistic Regression in the current study. The next step is to evaluate the relative importance of each risk factor in the ML-based predictive model. While regression analysis provides the coefficient and significant test results of each independent variable, the method of calculating the relative importance of each predicting factor to the outcome variable varies by the classification algorithms. When it comes to the Random Forests model, the contribution of each predicting factor can be determined by calculating the increases of the Gini Index when shuffling the factor. As explained above, a measure of impurity, the Gini Index, is used when making the decision about which predictor to split at each node. If one shuffles an important predictor, the impurity for all nodes should increase because the ability to build trees without the valuable information of the predictor disappears. The average decreases of the Gini Index by replacing the shuffled predictor with the original predictor can be calculated for each predictor. This measure is called the mean decreases of Gini Index. A higher score of the mean decreases of Gini Index indicates that the predictor is more important than other predictors.

Figure 11 represents the results of the mean decreases of Gini Index for each predictor that is used to build the Random Forests. As can be seen in Figure 11, the Maternal Relationship is the most important predictor of LCP offenders. Moreover, all familial risk factors except for Maternal Adolescent Pregnancy are proved to be the most important risk factors. In addition, the environmental risk factors are found to be important than individual risk factors. For instance, among the top 10 most important predictors, Delinquent Peers, Physical Environment, and School Environment are more important than Emotional Problem, Belief in Chance, and Use of substance. More importantly, the individual risk factors such as Race, Sex, Strain, and early onset of criminal involvement are not substantially important than familial and environmental risk factors.

Since the methods of evaluating the relative importance of risk factors between logistic regression and ML-based analyses are different, there is no absolute method to compare the importance of each factor between all predictive models. However, it is worth examining the standardized coefficients of variables in the logistic regression analysis to capture the differences. Table 27 presents the odds ratio and standardized coefficients of all independent variables of logistic regression analysis. Unlike the Random Forests model, logistic regression analysis is sensitive to the multicollinearity between independent variables, therefore, the coefficient should be interpreted with some caution. As can be seen in Table 27, four risk factors are found to be significantly associated with the variance of the dependent variable after controlling for the effects of other factors. The standardized coefficient of sex is the highest followed by gang exposure, risky neighborhood environments, and maternal adolescent pregnancy. To be specific, being male increases the likelihood of becoming LCP offenders by more than two times. Exposure to gang members increases the odds of becoming LCP offenders. Conversely, neighborhood environmental risks and maternal adolescent pregnancy decreases the likelihood of becoming LCP offenders.

Figure 11

Variable Importance Plot of Random Forests Model

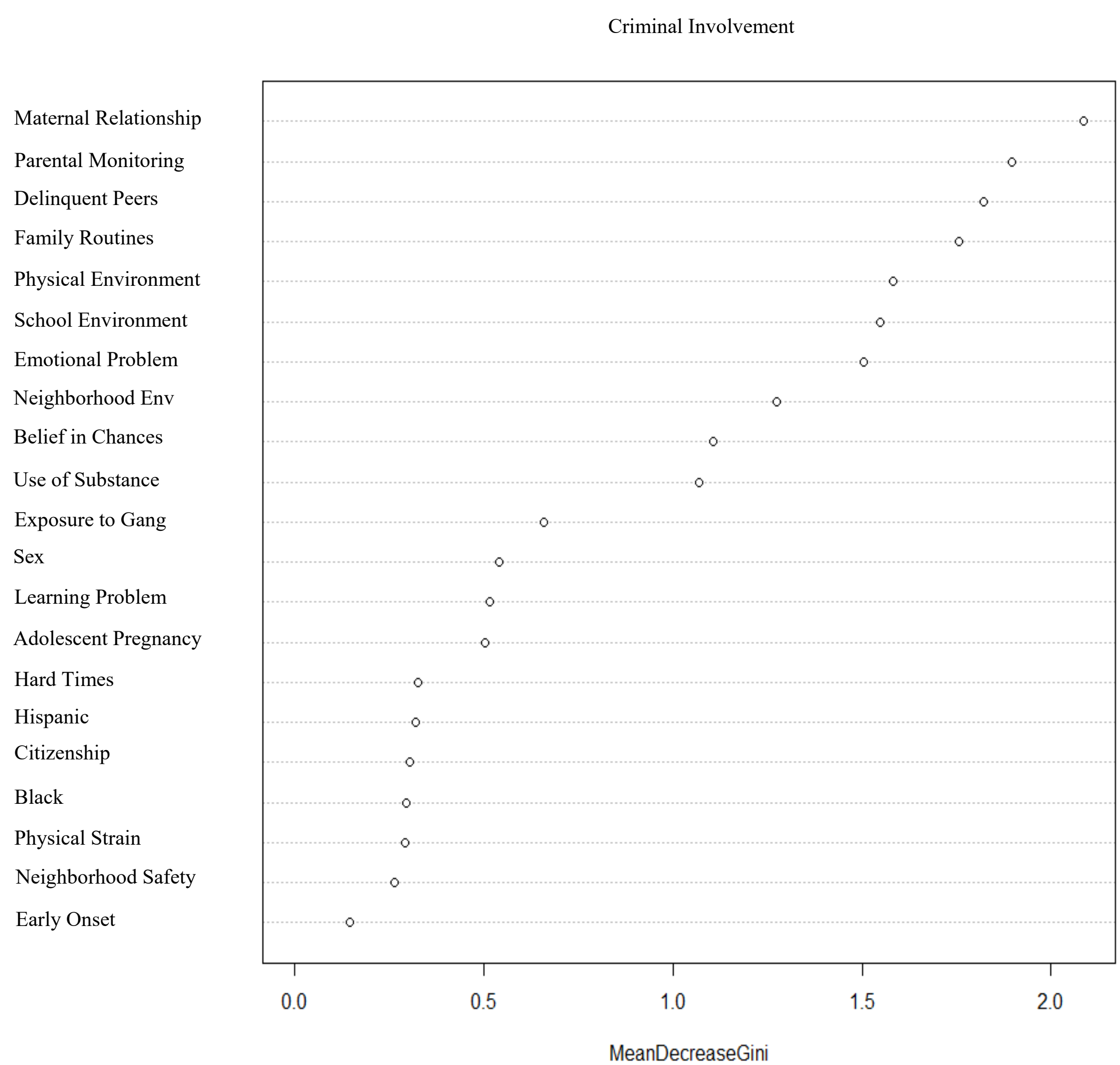


Table 27

Standardized Coefficients of Logistic Regression Model

	Odds Ratio (Standardized B)	SE	*p*
Family Factor			
Maternal Relationship	1.00 (.00)	.01	.94
Family Routines	1.00 (.00)	.02	.97
Parental Monitoring	.98 (-.08)	.02	.16
Maternal Adolescent Pregnancy	.52 (-.11)*	.16	.03
Individual Factor			
Emotional Problem	1.05 (.04)	.06	.40
Early Onset	1.18 (.02)	.45	.67
Belief in Chances of Getting Arrested	1.00 (.00)	.09	.99
Hard Times	.38 (-.11)	.19	.06
Physical/Emotional Strain	.88 (-.02)	.34	.75
Learning/Emotional Problem	1.00 (.00)	.30	.99
Use of Substance	1.13 (.07)	.10	.15
Citizenship	1.07 (.02)	.25	.76
Black	.94 (-.01)	.23	.78
Hispanic	1.23 (.04)	.31	.40
Sex (male)	2.21 (.19) ***	.10	.00
Peer and Environmental			
Delinquent Peers	1.01 (.04)	.02	.44
Exposure to Gang	1.93 (.17)***	.37	.00
Housing Environment	1.18 (.10)	.13	.13
Neighborhood Environment	.58 (-.17)*	.13	.02
School Environment	1.02 (.03)	.04	.60

(continued)

Environmental Safety	.53 (-.10)	.18	.06

CHAPTER V

Conclusion

Certainly, it is better to prevent crime before it happened than to recover the victim of crime and rehabilitate the criminals. Although the violent crime rates of the U. S. dramatically dropped in the past few decades, criminal incidents are still more frequently observed in the U.S. than in other developed countries (see Blumstein & Rosenfeld, 2008). A variety of efforts have been made to decrease crime rates, but numerous social problems were accompanied by the efforts. For instance, in 2010 the number of prisoners in the prison increased by three times than in the 1990s, therefore, social disorganization of disadvantaged neighborhoods and other damages to the minority communities were accompanied by the mass incarceration (Clear, 2009; Sampson & Loeffler, 2010; DeFina & Hannon, 2009). In addition, the incident history-based preventive policing scheme such as hot-spot policing successfully reduced crime (Braga, Papachristos, & Hureau, 2014), however, the policing tactic inevitably results in discriminating against people in disadvantaged neighborhoods (Rosenbaum, 2006).

In contrast to crime prevention tactics mentioned above, crime prevention through the early intervention focusing on high-risk individuals is less harmful to communities and individuals of disadvantaged neighborhoods, while it is proven to be more effective and cost-efficient than other crime prevention tactics (Farrington & Welsh, 2007; see Tremblay & Craig, 1995). At this instant, identifying the risky individuals who are more likely to become persistent criminal offenders is the key to improve the efficiency of early intervention programs.

Despite the fact that several risk assessment tools were developed and widely applied, the methods of identifying individuals at risk are still outdated to some extent. In most cases, risk scores of the assessment tools are summed to calculate a raw risk score, then individuals are categorized into high-risk or low-risk groups based on the raw risk score (see Schwalbe, 2007). Meanwhile, classification techniques have been significantly developed in the past few years with the development of Machine Learning techniques. Accordingly, empirical studies attempted to apply the ML-based predictive modeling to forecast recidivism (Tollenaar & Van der Heijden, 2013), and justice systems have begun to use ML-based forecasts when making parole release decisions (Berk, 2017).

To fill the gap, the current study aims to examine the predictive ability of ML techniques to identify life-course-persistent offenders. Along with the conventional statistical analysis, Logistic Regression, six ML-based non-parametric classification techniques are applied to build the predictive modeling. Using a nationally represented sample drawn from the National Longitudinal Survey of Youth 1997, LCP offenders are forecasted by risk factors rooted in a variety of criminological theories and empirical evidence.

Before building the predictive modeling, the LCP and AL offenders are identified by using Latent Class Growth Analysis. Trajectories of individual criminal involvement from the age of 14 to 19 are examined to find the distinct classes comprising similar patterns of delinquent behaviors over time. According to the LCGA, the model with three latent classes is found to be best fitted to the data. As a result of the class identification, a majority of individuals (n = 4,555) are identified as a low-offense group.

In line with the previous studies of identifying LCP and AL offenders using the Latent Growth Modeling approach, approximately 12% (n = 676) of individuals are identified as AL offenders, while 3% (n = 188) of individuals are classified into the LCP group (Nagin, 1999).

Upon the review of all questions in the NLSY97 survey, 21 variables within three categories are selected to build the predicting models. The first dimension of risk factors is family-related factors. Maternal relationships, family routines, parental monitoring, and maternal adolescent pregnancy are included in the family-related factor. The second dimension of risk factors is individual risk factors. Emotional problems, early onset of crime, belief in chances of getting arrested, hard time experience, physical/emotional strain, learning/emotional problem, use of substance, citizenship, race, and sex are included in the individual risk factor. The third dimension is comprised of peer and environmental risk factors. Delinquent peers, exposure to gang members, housing environment, neighborhood environment, school environment, and environmental safety are included in the peer and environmental risk factors.

Does the Predictive Approach Work?

The first question that should be answered by this study is whether the prediction of Life-Course-Persistent offenders using parametric and non-parametric statistical analysis is achievable. To administrators and decision-makers in juvenile justice systems, the implementation of more comprehensive and standardized risk assessment tools is demanded in estimating the future risk of youths and adolescents (see Vincent, Gui, & Grisso, 2012). Given that the standardized risk assessment tools are going to be constantly used and the data from the assessment are massively accumulated, the stakeholders in the juvenile justice system should be sought for more accurate forecasting methodologies. To answer the first research question, seven predictive models using Regression analysis and Machine Learning are built for the classification between LCP and AL offenders.

First, for the baseline model, the prediction of LCP using Logistic Regression analysis is conducted. Results of the prediction show that the overall accuracy of the LR model is relatively high (=.80), although only 27% of actual LCP offenders are correctly forecasted. That being said, LR based predictive model missed 73% of actual LCP offenders. Secondly, the overall accuracy of the CART (*tree*) model is slightly improved than LR by 2 percent (= .82). However, the DT model also misclassified a large number of LCP offenders, thus, 69 percent of actual LCP offenders are misclassified. Thirdly, the CART (*rpart*) model also yields good overall accuracy (=.78), however, the model forecasts only 15 percent of actual LCP offenders correctly. Fourthly, the Naïve Bayes model showed different classification patterns comparing to the previous three models. Although the overall accuracy of the Naïve Bayes model decreases (=.67), the model misses only 40 percent of the actual LCP offenders. Fifthly, the balanced Random Forests model provides similar results to those of the Naïve Bayes model. The overall accuracy of RF is decent (= .67), while 59 percent of actual LCP offenders are correctly forecasted as LCP offenders by the RF model. Sixthly, the single-layered Neural Networks model with three perceptrons also provides similar levels of predictive performance to those of Naïve Bayes and Random Forests models. While the overall accuracy slightly decreased (=.65), the hit rate of LCP offenders increased (=.61). Lastly, the Deep Learning model using the *h2o* package predicted 69% of actual LCP offenders correctly, while the overall accuracy (=.69) is higher than that of Naïve Bayes, Random Forests, and the single-layered NNs model. In conclusion. fine-tuned ML-based forecasting models such as balanced RF, NNs, and Deep Learning can successfully predict the LCP offenders using theory-based risk assessment tools.

More importantly, the current study revealed that familial and environmental risk factors are more important than individual risk factors for the prediction of LCP offenders. As can be seen in Figure 11, most individual risk factors, such as Race, Sex, and Strain are less important than familial risk factors and environmental risk factors. Although the Machine Learning based predictive model is criticized for its Black Box feature, the relative importance of each variable may provide useful information to decision-makers and stakeholders of the juvenile justice system.

ML better than Logistic Regression in Predicting LCP Offenders?

The second research question is to answer whether ML-based predictive models outperform the conventional statistical analysis Logistic Regression in predicting LCP offenders. As can be seen in the summary table, none of the performance parameters of the LR model is the highest compared to those of the ML-based model respectively. Bearing in mind that the overall accuracy is not the main concern in predicting chronic offenders, the sensitivity and AUC should be examined among all predictive models.

When it comes to sensitivity, also called hit rate, the Deep Learning model performed best. In other words, if administrators or decision-makers seek out for identifying as many high-risk juveniles as possible who would actually become persistent offenders, the Deep Learning model would be the most suitable method. However, there still are concerns about cost efficiency when using Deep Learning for the prediction of consistent offenders. The predictive model forecasted that 215 individuals would be AL offenders while 130 individuals would be LCP offenders. Out of the 130 individuals classified into LCP offenders, only 38 percent (49 out of 130) actually become LCP offenders. Given the restricted resources of the juvenile justice system, the cost and efforts of an intervention program to those who were labeled as LCP but not actually become LCP offenders may be a substantial waste. Therefore, the administrators and practitioners are required to determine the most suitable cost ratio under the circumstances of their agency.

To determine the overall predictive performance, Area Under the Curve (AUC) should be examined. AUC tells how well the model predicts zeros as zeros and ones as ones. Being said, the higher the AUC, the better the model is at separating LCP and AL offenders. Despite the poor sensitivity, Logistic Regression model (= .708) shows a relatively high AUC among seven predictive models. However, the AUC is still lower than that of the Random Forests (=.717) model.

In conclusion, the prediction of LCP using Logistic Regression performed as well as some conventional Machine Learning models. However, it should be remarked that advanced Machine Learning techniques such as balanced Random Forests and Deep Learning yield better predicting performance. Further, both Machine Learning techniques are flexible to adjust the ratio between false positives and false negatives. For that reason, it can be argued that the machine learning models can provide more valuable information regarding the future risk of juveniles to assist decision-makers of juvenile justice systems.

Implementation

In recent years, criminal justice agencies in a few states began to adopt assistance from machine learning-based predictive modeling in estimating the future risk of offenders. In contrast, from the best of my knowledge, juvenile justice systems have yet considered applying the ML-based predictive models, and only a few pieces of research have examined the predictive performance of ML-based forecasting of high-risk offenders (Kim et al., 2019). Using nationally representative longitudinal samples in the U.S., the current study attempts to predict LCP offenders based on several Machine Learning techniques.

Although the risk factors used in this study are not drawn from standardized risk assessment tools such as COMPAS and SAVRY, ML algorithms provide moderate performance in identifying LCP offenders. AUCs for the Random Forests and Deep Learning model show the low .70s, while AUC for the other ML-based models range from the low .60s to mid .60s. Moreover, the sensitivity scores, the rate of classifying actual LCPs as LCPs, of the Random Forests, Neural Networks, and Deep Learning range from the high .50s to the mid .60s. Taken all performance metrics together, it is concluded that Random Forests and Deep Learning algorithms provide the acceptable performance of predicting LCP offenders. Consequently, it is reasonable to argue that practitioners and administrators of juvenile justice should consider implementing Machine Learning algorithms for estimating the future risk of juvenile offenders.

Limitation

Despite the valuable findings of this study, the results should be interpreted with some caution. First, the risk assessment tools used for this study are not the standardized form that is widely applied in estimating the future risk of juveniles. Due to data access restrictions, the current study utilizes publicly available longitudinal data and finds the most plausible risk factors based on theoretical backgrounds among thousands of items in the survey. The use of unstandardized risk assessment tools may result in the moderate performance of prediction. Second, the small size of the sample for the analysis makes this study unavailable to apply cross-validation techniques such as k-fold cross-validation and bootstrapping. In other words, the forecasting results may be changed if the training and testing data would be assigned differently. Third, AUC of the predictive models in this study is low when comparing to other studies predicting adult recidivism (Na, Song, Oh, & Park, 2021; Ozkan, 2017; Berk et al., 2012; Brennan, Dieterich, & Ehret, 2009;). However, low AUC for the predictive model of juvenile offenders is also observed in previous studies using raw scores or weighted scores of risk assessment instruments for the classification (see Schwalbe, 2007). For instance, studies using standardized risk assessment tools for the prediction of juvenile recidivism reported that the AUCs of the predictive model range from high .50s to low .70s (Barnoski, 2004; Johnson, Wagner, & Mathews, 2002; Krysik & LeCroy, 2002). This implies that the current actuarial risk assessment tools relying on the environmental and behavioral risk factors hit the limit for improving its predictive accuracy. This limit has been pointed out by several researchers (Yang, Wong, & Coid, 2010; Poldrack et al., 2017), and they argued that the more advanced risk instruments such as biologically based factors should be included (Poldrack et al., 2017).

REFERENCES

Agnew, R. (1992). Foundation for a general strain theory of crime and delinquency. *Criminology*, *30*(1), 47–88.

Agnew, R., Brezina, T., Wright, J. P., & Cullen, F. T. (2002a). Strain, personality traits, and delinquency: Extending general strain theory. *Criminology*, *40*(1), 43–72.

Agnew, R., Brezina, T., Wright, J. P., & Cullen, F. T. (2002b). Strain, personality traits, and delinquency: Extending general strain theory. *Criminology*, *40*(1), 43–72.

Agnew, R., & White, H. R. (1992). An empirical test of general strain theory. *Criminology*, *30*(4), 475–500.

Akers, R. L., & Silverman, A. (2004). Toward a social learning model of violence and terrorism. *Violence: From theory to research*, 19–35.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, *16*(7), 16–07.

Andrews, D. A., Bonta, J., & Wormith, S. J. (2000). *Level of service/case management inventory: LS/CMI*. Multi-Health Systems Toronto, Canada.

Aseltine Jr, R. H., Gore, S., & Gordon, J. (2000). Life stress, anger and anxiety, and delinquency: An empirical test of general strain theory. *Journal of health and social behavior*, 256–275.

Auer, P., Cesa-Bianchi, N., Freund, Y., & Schapire, R. E. (1995). Gambling in a rigged casino: The adversarial multi-armed bandit problem. *Proceedings of IEEE 36th Annual Foundations of Computer Science*, 322–331.

Barnoski, R. (2004). Assessing risk for re-offense: Validating the Washington State Juvenile Court assessment appendices. *Olympia, WA: Washington State Institute for Public Policy*, 1–49.

Beaver, K. M., & Wright, J. P. (2007). A child effects explanation for the association between family risk and involvement in an antisocial lifestyle. *Journal of Adolescent Research*, *22*(6), 640–664.

Bergman, L. R., & Andershed, A.-K. (2009). Predictors and outcomes of persistent or age-limited registered criminal behavior: A 30-year longitudinal study of a Swedish urban population. *Aggressive Behavior*, *35*(2), 164–178. https://doi.org/10.1002/ab.20298

Berk, R. (2017). An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. *Journal of Experimental Criminology*, *13*(2), 193–216.

Berk, R. A. (2008). *Statistical learning from a regression perspective* (Vol 14). Springer.

Berk, R. A., Kriegler, B., & Baek, J.-H. (2006a). Forecasting dangerous inmate misconduct: An application of ensemble statistical procedures. *Journal of Quantitative Criminology*, *22*(2), 131–145.

Berk, R. A., Kriegler, B., & Baek, J.-H. (2006b). Forecasting dangerous inmate misconduct: An application of ensemble statistical procedures. *Journal of Quantitative Criminology*, *22*(2), 131–145.

Berk, R., & Bleich, J. (2014). Forecasts of violence to inform sentencing decisions. *Journal of Quantitative Criminology*, *30*(1), 79–96.

Berk, R., Li, A., & Hickman, L. J. (2005a). Statistical difficulties in determining the role of race in capital cases: A re-analysis of data from the state of Maryland. *Journal of Quantitative Criminology*, *21*(4), 365–390.

Berk, R., Li, A., & Hickman, L. J. (2005b). Statistical difficulties in determining the role of race in capital cases: A re-analysis of data from the state of Maryland. *Journal of Quantitative Criminology*, *21*(4), 365–390.

Berk, R., Sherman, L., Barnes, G., Kurtz, E., & Ahlman, L. (2009a). Forecasting murder within a population of probationers and parolees: A high stakes application of statistical learning. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(1), 191–211.

Berk, R., Sherman, L., Barnes, G., Kurtz, E., & Ahlman, L. (2009b). Forecasting murder within a population of probationers and parolees: A high stakes application of statistical learning. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(1), 191–211.

Blair, J., Mitchell, D., & Blair, K. (2005). *The psychopath: Emotion and the brain.* Blackwell Publishing.

Blumstein, A., & Rosenfeld, R. (2008). Factors contributing to US crime trends. *Understanding crime trends: Workshop report*, *2*, 13–44.

Booth, J. A., Farrell, A., & Varano, S. P. (2008a). Social control, serious delinquency, and risky behavior: A gendered analysis. *Crime & Delinquency*, *54*(3), 423–456.

Booth, J. A., Farrell, A., & Varano, S. P. (2008b). Social control, serious delinquency, and risky behavior: A gendered analysis. *Crime & Delinquency*, *54*(3), 423–456.

Braga, A. A., Papachristos, A. V., & Hureau, D. M. (2014). The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice quarterly*, *31*(4), 633–663.

Breiman, L. (1998). Arcing classifier (with discussion and a rejoinder by the author). *The annals of statistics*, *26*(3), 801–849.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Brennan, T., Dieterich, W., & Ehret, B. (2009). Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Criminal Justice and Behavior*, *36*(1), 21–40.

Broidy, L. M. (2001). A test of general strain theory. *Criminology*, *39*(1), 9–36.

Burgess, E. W. (1928). Residential segregation in American cities. *The annals of the american academy of political and social science*, *140*(1), 105–115.

Burgess, R. L., & Akers, R. L. (1966). A differential association-reinforcement theory of criminal behavior. *Social problems*, *14*(2), 128–147.

Burscher, B., Vliegenthart, R., & De Vreese, C. H. (2015). Using supervised machine learning to code policy issues: Can classifiers generalize across contexts? *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 122–131.

Burt, C. H., Simons, R. L., & Simons, L. G. (2006). A longitudinal test of the effects of parenting and the stability of self-control: Negative evidence for the general theory of crime. *Criminology*, *44*(2), 353–396.

Bushway, S. D., Sweeten, G., & Wilson, D. B. (2006). Size matters: Standard errors in the application of null hypothesis significance testing in criminology and criminal justice. *Journal of Experimental Criminology*, *2*(1), 1–22.

Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. *Nature*, *559*(7715), 547–555.

Caplan, J. M., Kennedy, L. W., & Miller, J. (2011). Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting. *Justice quarterly*, *28*(2), 360–381.

Caulkins, J., Cohen, J., Gorr, W., & Wei, J. (1996a). Predicting criminal recidivism: A comparison of neural network models with statistical methods. *Journal of Criminal Justice*, *24*(3), 227–240.

Caulkins, J., Cohen, J., Gorr, W., & Wei, J. (1996b). Predicting criminal recidivism: A comparison of neural network models with statistical methods. *Journal of Criminal Justice*, *24*(3), 227–240.

Cernkovich, S. A., & Giordano, P. C. (1987). Family relationships and delinquency. *Criminology*, *25*(2), 295–319.

Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley*, *110*(1–12), 24.

Clear, T. R. (2009). *Imprisoning communities: How mass incarceration makes disadvantaged neighborhoods worse*. Oxford University Press.

Cohen, L. E., Kluegel, J. R., & Land, K. C. (1981). Social inequality and predatory criminal victimization: An exposition and test of a formal theory. *American sociological review*, 505–524.

Coid, J., Yang, M., Ullrich, S., Zhang, T., Roberts, A., Roberts, C., Rogers, R., & Farrington, D. (2007). Predicting and understanding risk of re-offending: The Prisoner Cohort Study. *Research summary*, *6*(1), 9.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American psychologist*, *34*(7), 571.

DeFina, R., & Hannon, L. (2013). The impact of mass incarceration on poverty. *Crime & Delinquency*, *59*(4), 562–586.

Derogatis, L. R., & Melisaratos, N. (1983a). The brief symptom inventory: An introductory report. *Psychological medicine*, *13*(3), 595–605.

Derogatis, L. R., & Melisaratos, N. (1983b). The brief symptom inventory: An introductory report. *Psychological medicine*, *13*(3), 595–605.

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and psychopathology*, *18*(3), 791.

Dubin, R. (1978). *Theory Building* (2nd edition). Free Pr.

Dubow, E. F., Huesmann, L. R., Boxer, P., & Smith, C. (2014). Childhood predictors and age 48 outcomes of self-reports and official records of offending. *Criminal Behaviour and Mental Health*, *24*(4), 291–304.

Duwe, G., & Kim, K. (2017). Out with the old and in with the new? An empirical comparison of supervised learning algorithms to predict recidivism. *Criminal Justice Policy Review*, *28*(6), 570–600.

Duwe, G., & Kim, K. (2018). The Neglected "R" in the Risk-Needs-Responsivity Model: A New Approach for Assessing Responsivity to Correctional Interventions. *Justice Evaluation Journal*, *1*(2), 130–150.

Farrington, D. P. (2000). *Adolescent violence: Findings and implications from the Cambridge Study.*

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: Systematic review and meta-analysis. *Bmj*, *345*, e4692.

Gardner, W., Lidz, C. W., Mulvey, E. P., & Shaw, E. C. (1996). A comparison of actuarial methods for identifying repetitively violent patients with mental illnesses. *Law and Human Behavior*, *20*(1), 35–48.

Giordano, P. C., Cernkovich, S. A., & Rudolph, J. L. (2002). Gender, crime, and desistance: Toward a theory of cognitive transformation. *American journal of sociology*, *107*(4), 990–1064.

Gottfredson, D. M. (1987). Prediction and classification in criminal justice decision making. *Crime and Justice*, *9*, 1–20.

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

Gottfredson, S. D., & Gottfredson, D. M. (1980). Screening for risk: A comparison of methods. *Criminal Justice and Behavior*, *7*(3), 315–330.

Grann, M., & Långström, N. (2007). Actuarial assessment of violence risk: To weigh or not to weigh? *Criminal Justice and Behavior*, *34*(1), 22–36.

Hamilton, Z., Campbell, C. M., van Wormer, J., Kigerl, A., & Posey, B. (2016). Impact of swift and certain sanctions: Evaluation of Washington State's policy for offenders on community supervision. *Criminology & Public Policy*, *15*(4), 1009–1072.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.

Hay, C., & Forrest, W. (2006). The development of self-control: Examining self-control theory's stability thesis. *Criminology*, *44*(4), 739–774.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of science*, *15*(2), 135–175.

Hinkle, J. C., Weisburd, D., Telep, C. W., & Petersen, K. (2020). Problem-oriented policing for reducing crime and disorder: An updated systematic review and meta-analysis. *Campbell Systematic Reviews*, *16*(2), e1089.

Hirschi, T. (1969). A control theory of delinquency. *Criminology theory: Selected classic readings*, *1969*, 289–305.

Hoeve, M., Stams, G. J. J., Van der Put, C. E., Dubas, J. S., Van der Laan, P. H., & Gerris, J. R. (2012). A meta-analysis of attachment to parents and delinquency. *Journal of abnormal child psychology*, *40*(5), 771–785.

Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science*, *355*(6324), 486–488.

Holleran, D., & Stout, B. D. (2017a). The importance of race in juvenile commitment in the New Jersey Family Court. *Crime & Delinquency*, *63*(3), 353–372.

Holleran, D., & Stout, B. D. (2017b). The importance of race in juvenile commitment in the New Jersey Family Court. *Crime & Delinquency*, *63*(3), 353–372.

Holloway, K., Bennett, T., & Farrington, D. P. (2008). Crime prevention research review no. 3: Does Neighborhood Watch reduce crime. *Washington, DC: US Department of Justice Office of Community Oriented Policing Services*.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. Wiley New York.

Huizinga, D., & Elliott, D. S. (1987). Juvenile offenders: Prevalence, offender incidence, and arrest rates by race. *Crime & Delinquency*, *33*(2), 206–223.

Hunt, P., Hollywood, J. S., & Saunders, J. M. (2014). *Evaluation of the Shreveport predictive policing experiment*. Rand Corporation Santa Monica.

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The annals of applied statistics*, *2*(3), 841–860.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Support vector machines. In *An Introduction to Statistical Learning* (pp 337–372). Springer.

Jang, S. J., & Johnson, B. R. (2003). Strain, negative emotions, and deviant coping among African Americans: A test of general strain theory. *Journal of quantitative criminology*, *19*(1), 79–105.

Jang, S. J., & Smith, C. A. (1997). A test of reciprocal causal relationships among parental supervision, affective ties, and delinquency. *Journal of Research in Crime and Delinquency*, *34*(3), 307–336.

Jeffery, C. R. (1971). *Crime prevention through environmental design* (Vol 91). Sage publications Beverly Hills, CA.

Johnson, K., Wagner, D., & Matthews, T. (2002). Missouri juvenile risk assessment re-validation report. *Madison, WI: National Council on Crime and Delinquency*.

Jolliffe, D., & Farrington, D. P. (2009). A systematic review of the relationship between childhood impulsiveness and later violence. *Personality, personality disorder, and violence*, 41–61.

Jolliffe, D., Farrington, D. P., Piquero, A. R., Loeber, R., & Hill, K. G. (2017). Systematic review of early risk factors for life-course-persistent, adolescence-limited, and late-onset offenders in prospective longitudinal studies. *Aggression and violent behavior*, *33*, 15–23.

Jolliffe, D., Farrington, D. P., Piquero, A. R., MacLeod, J. F., & Van de Weijer, S. (2017). Prevalence of life-course-persistent, adolescence-limited, and late-onset offenders: A systematic review of prospective longitudinal studies. *Aggression and violent behavior*, *33*, 4–14.

Jolliffe, D., Murray, J., Farrington, D., & Vannick, C. (2012). Testing the Cambridge Quality Checklists on a review of disrupted families and crime. *Criminal Behaviour and Mental Health*, *22*(5), 303–314.

Jones, B. L., Nagin, D. S., & Roeder, K. (2001). A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological methods & research*, *29*(3), 374–393.

Jung, T., & Wickrama, K. A. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and personality psychology compass*, *2*(1), 302–317.

Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhyā: The Indian Journal of Statistics, Series A*, 49–66.

Kim, K., Duwe, G., Tiry, E., Oglesby-Neal, A., Hu, C., Shields, R., Letourneau, E., Caldwell, M., Institute, T. U., & America, U. S. of. (2019). *Development and Validation of an Actuarial Risk Assessment Tool for Juveniles with a History of Sexual Offending*. Office of Justice Programs, National Criminal Justice Reference Service.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 2053951714528481. https://doi.org/10.1177/2053951714528481

Kort-Butler, L. A. (2010). Experienced and vicarious victimization: Do social support and self-esteem prevent delinquent responses? *Journal of Criminal Justice*, *38*(4), 496–505.

Krysik, J., & LeCroy, C. W. (2002). The empirical validation of an instrument to predict risk of recidivism among juvenile offenders. *Research on Social Work Practice*, *12*(1), 71–81.

Lachenbruch, P. A. (2014). McNemar test. *Wiley StatsRef: Statistics Reference Online*.

Laub, J. H., & Sampson, R. J. (1988). Unraveling families and delinquency: A reanalysis of the gluecks' data. *Criminology*, *26*(3), 355–380.

Leffondré, K., Abrahamowicz, M., Regeasse, A., Hawker, G. A., Badley, E. M., McCusker, J., & Belzile, E. (2004). Statistical measures were proposed for identifying longitudinal patterns of change in quantitative health indicators. *Journal of clinical epidemiology*, *57*(10), 1049–1062.

Levine, E. S., Tisch, J., Tasso, A., & Joy, M. (2017). The New York City police department's domain awareness system. *Interfaces*, *47*(1), 70–84.

Liaw, A., & Weiner, M. (2002). *Classification and Regression by randomForest. R News (2/3). Retrieved on April 7, 2012*.

Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, *16*(6), 321–332.

Lidz, C. W., Mulvey, E. P., & Gardner, W. (1993). The accuracy of predictions of violence to others. *Jama*, *269*(8), 1007–1011.

Loeber, R., & Stouthamer-Loeber, M. (1986). Family factors as correlates and predictors of juvenile conduct problems and delinquency. *Crime and justice*, *7*, 29–149.

Loh, W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 14–23.

Lohr, S. (2012). The age of big data. *New York Times*, *11*(2012).

Matsueda, R. L., & Heimer, K. (1997). A symbolic interactionist theory of role-transitions, role-commitments, and delinquency. *Developmental theories of crime and delinquency*, *44*(3), 163–213.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, *27*(4), 12–12.

Meijer, A., & Wessels, M. (2019). Predictive policing: Review of benefits and drawbacks. *International Journal of Public Administration*, *42*(12), 1031–1039.

Minsky, M., & Papert, S. (1969). An introduction to computational geometry. *Cambridge tiass., HIT*.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological review*, *100*(4), 674.

Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and psychopathology*, *13*(2), 355–375.

Moffitt, T. E., Caspi, A., Dickson, N., Silva, P., & Stanton, W. (1996). Childhood-onset versus adolescent-onset antisocial conduct problems in males: Natural history from ages 3 to 18 years. *Development and psychopathology*, *8*(2), 399–424.

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.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of consulting and clinical psychology*, *62*(4), 783.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological methods*, *4*(2), 139.

Neuilly, M.-A., Zgoba, K. M., Tita, G. E., & Lee, S. S. (2011). Predicting recidivism in homicide offenders using classification tree analysis. *Homicide studies*, *15*(2), 154–176.

Ngo, F. T., Govindu, R., & Agarwal, A. (2015). Assessing the predictive utility of logistic regression, classification and regression tree, chi-squared automatic interaction detection, and neural network models in predicting inmate misconduct. *American Journal of Criminal Justice*, *40*(1), 47–74.

Oh, G., & Connolly, E. J. (2019). Anger as a mediator between peer victimization and deviant behavior in South Korea: A cross-cultural application of general strain theory. *Crime & Delinquency*, *65*(8), 1102–1122.

Ozkan, T. (2017). *Predicting recidivism through machine learning (Doctoral Dissertation)*. The University of Texat at Dallas, Texas.

Ozkan, T., Clipper, S. J., Piquero, A. R., Baglivio, M., & Wolff, K. (2020). Predicting sexual recidivism. *Sexual Abuse*, *32*(4), 375–399.

Palocsay, S. W., Wang, P., & Brookshire, R. G. (2000). Predicting criminal recidivism using neural networks. *Socio-Economic Planning Sciences*, *34*(4), 271–284.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, *42*(4), 2162–2172.

Paternoster, R., Brame, R., Bacon, S., Ditchfield, A., Beckman, K., & Frederique, N. (2003). *An empirical analysis of Maryland's death sentencing system with respect to the influence of race and legal jurisdiction*. University of Maryland, College Park.

Petrosino, A., Petrosino, C. T., & Buehler, J. (2005). "Scared Straight" and other juvenile awareness programs for preventing juvenile delinquency. *Campbell Systematic Reviews*, *1*(1), 1–62.

Pflueger, M. O., Franke, I., Graf, M., & Hachtel, H. (2015). Predicting general criminal recidivism in mentally disordered offenders using a random forest approach. *BMC psychiatry*, *15*(1), 62.

Piquero, A. R., Farrington, D. P., Welsh, B. C., Tremblay, R., & Jennings, W. G. (2008). Effects of early family/parent training programs on antisocial behavior & delinquency. *Campbell Systematic Reviews*, *4*(1), 1–122.

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J.-B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature reviews neuroscience*, *18*(2), 115.

Poyner, B. (1993). What works in crime prevention: An overview of evaluations. *Crime prevention studies*, *1*, 7–34.

Pratt, T. C., & Cullen, F. T. (2000). The empirical status of Gottfredson and Hirschi's general theory of crime: A meta-analysis. *Criminology*, *38*(3), 931–964.

Pratt, T. C., Cullen, F. T., Blevins, K. R., Daigle, L., & Unnever, J. D. (2002). The relationship of attention deficit hyperactivity disorder to crime and delinquency: A meta-analysis. *International Journal of Police Science & Management*, *4*(4), 344–360.

Pratt, T. C., Cullen, F. T., Sellers, C. S., Thomas Winfree Jr, L., Madensen, T. D., Daigle, L. E., Fearn, N. E., & Gau, J. M. (2010). The empirical status of social learning theory: A meta-analysis. *Justice Quarterly*, *27*(6), 765–802.

Purtill, R. (1970). The purpose of science. *Philosophy of Science*, *37*(2), 301–306.

Raine, A. (1993). Features of borderline personality and violence. *Journal of Clinical Psychology*, *49*(2), 277–281.

Rankin, J. H., & Kern, R. (1994). Parental attachments and delinquency. *Criminology*, *32*(4), 495–515.

Razi, M. A., & Athappilly, K. (2005). A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications*, *29*(1), 65–74.

Rebellon, C. J., Manasse, M. E., Van Gundy, K. T., & Cohn, E. S. (2012). Perceived injustice and delinquency: A test of general strain theory. *Journal of Criminal Justice*, *40*(3), 230–237.

Rosenbaum, D. P. (2006). The limits of hot spots policing. *Police innovation: Contrasting perspectives*, 245–263.

Rosenfeld, B., & Lewis, C. (2005). Assessing violence risk in stalking cases: A regression tree approach. *Law and Human Behavior*, *29*(3), 343–357.

Rowe, D. C., & Farrington, D. P. (1997). The familial transmission of criminal convictions. *Criminology*, *35*(1), 177–202.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, *323*(6088), 533–536.

Rutter, M., Moffitt, T. E., & Caspi, A. (2006). Gene–environment interplay and psychopathology: Multiple varieties but real effects. *Journal of child Psychology and Psychiatry*, *47*(3-4), 226–261.

Sajda, P. (2006). Machine learning for detection and diagnosis of disease. *Annu. Rev. Biomed. Eng.*, *8*, 537–565.

Sampson, R. J., & Laub, J. H. (2003). Life-course desisters? Trajectories of crime among delinquent boys followed to age 70. *Criminology*, *41*(3), 555–592.

Sampson, R. J., & Laub, J. H. (2005). A life-course view of the development of crime. *The Annals of the American Academy of Political and Social Science*, *602*(1), 12–45.

Sampson, R. J., & Loeffler, C. (2010). Punishment's place: The local concentration of mass incarceration. *Daedalus*, *139*(3), 20–31.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, *3*(3), 210–229.

Schapire, R. E., Freund, Y., Bartlett, P., & Lee, W. S. (1998). Boosting the margin: A new explanation for the effectiveness of voting methods. *The annals of statistics*, *26*(5), 1651–1686.

Schwalbe, C. S. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law and human behavior*, *31*(5), 449.

Schwalbe, C. S., Fraser, M. W., Day, S. H., & Cooley, V. (2006). Classifying juvenile offenders according to risk of recidivism: Predictive validity, race/ethnicity, and gender. *Criminal justice and behavior*, *33*(3), 305–324.

Scriven, M. (1959). Explanation and prediction in evolutionary theory. *Science*, *130*(3374), 477–482.

Seni, G., & Elder, J. F. (2010). Ensemble methods in data mining: Improving accuracy through combining predictions. *Synthesis lectures on data mining and knowledge discovery*, *2*(1), 1–126.

Shah, D. V., Cappella, J. N., & Neuman, W. R. (2015). Big data, digital media, and computational social science: Possibilities and perils. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 6–13.

Shmueli, G. (2010). To explain or to predict? *Statistical science*, *25*(3), 289–310.

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS quarterly*, 553–572.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., & Bolton, A. (2017). Mastering the game of go without human knowledge. *nature*, *550*(7676), 354–359.

Silver, E., Smith, W. R., & Banks, S. (2000). Constructing actuarial devices for predicting recidivism: A comparison of methods. *Criminal Justice and behavior*, *27*(6), 733–764.

Skinner, B. F. (2019). *The behavior of organisms: An experimental analysis*. BF Skinner Foundation.

Skinner, W. F., & Fream, A. M. (1997). A social learning theory analysis of computer crime among college students. *Journal of research in crime and delinquency*, *34*(4), 495–518.

Steadman, H. J., Silver, E., Monahan, J., Appelbaum, P., Robbins, P. C., Mulvey, E. P., Grisso, T., Roth, L. H., & Banks, S. (2000). A classification tree approach to the development of actuarial violence risk assessment tools. *Law and human behavior*, *24*(1), 83–100.

Sugiyama, M. (2015). *Introduction to statistical machine learning*. Morgan Kaufmann.

Sutherland, E. (1956). *'A Sociological Theory of Criminal Behavior*.

Sylvestre, M.-P., Vatnik, D., & Vatnik, M. D. (2014). *Package 'traj'*.

Thornberry, T. P. (1987). Toward an interactional theory of delinquency. *Criminology*, *25*(4), 863–892.

Thornberry, T. P., Giordano, P. C., Uggen, C., Matsuda, M., Masten, A. S., Bulten, E., & Donker, A. G. (2012). Explanations for offending. *From Juvenile delinquency to adult crime: Criminal careers, justice policy and prevention*, 47–85.

Thornberry, T. P., & Krohn, M. D. (2005). Applying interactional theory to the explanation of continuity and change in antisocial behavior. *Integrated developmental and life-course theories of offending*, *14*, 183–209.

Tolan, S., Miron, M., Gómez, E., & Castillo, C. (2019). Why machine learning may lead to unfairness: Evidence from risk assessment for juvenile justice in catalonia. *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, 83–92.

Tollenaar, N., & Van der Heijden, P. G. M. (2013). Which method predicts recidivism best?: A comparison of statistical, machine learning and data mining predictive models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(2), 565–584.

Tremblay, R. E., & Craig, W. M. (1995). Developmental crime prevention. *Crime and justice*, *19*, 151–236.

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, *49*(11), 1225–1231.

Uchida, C. D. (2009). *A national discussion on predictive policing: Defining our terms and mapping successful implementation strategies*. National Institute of Justice Los Angeles.

Vincent, G. M., Guy, L. S., & Grisso, T. (2012). *Risk assessment in juvenile justice: A guidebook for implementation*.

Warner, B., & Misra, M. (1996). Understanding neural networks as statistical tools. *The american statistician*, *50*(4), 284–293.

Warr, M., & Stafford, M. (1991). The influence of delinquent peers: What they think or what they do? *Criminology*, *29*(4), 851–866.

Weber, M. (1978). *Economy and society: An outline of interpretive sociology* (Vol 1). Univ of California Press.

Wiatrowski, M. D., Griswold, D. B., & Roberts, M. K. (1981). Social control theory and delinquency. *American sociological review*, 525–541.

Wright, B. R. E., Caspi, A., Moffitt, T. E., & Silva, P. A. (1999). Low self-control, social bonds, and crime: Social causation, social selection, or both? *Criminology*, *37*(3), 479–514.

Yang, M., Liu, Y., & Coid, J. (2010). Applying Neural Networks and other statistical models to the classification of serious offenders and the prediction of recidivism. *Ministry of Justice Research Series*, *6*(10).

Yang, M., Wong, S. C., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological bulletin*, *136*(5), 740.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122.

Zamith, R., & Lewis, S. C. (2015). Content analysis and the algorithmic coder: What computational social science means for traditional modes of media analysis. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 307–318.

Zara, G., & Farrington, D. P. (2013). Assessment of risk for juvenile compared with adult criminal onset implications for policy, prevention, and intervention. *Psychology, Public Policy, and Law*, *19*(2), 235.

Zara, G., & Farrington, D. P. (2020). Childhood Risk Factors for Self-Reported Versus Official Life-Course-Persistent, Adolescence-Limited, and Late-Onset Offending. *Criminal Justice and Behavior*, *47*(3), 352–368.

VITA

Gyeongseok Oh

EDUCATION

2017-2021	Ph.D., *Criminal Justice*, Sam Houston State University Dissertation: *Predicting life-course persistent offending using machine learning*
2013-2016	M.S., *Criminology and Criminal Justice*, Florida State University
2011-2013	M.A., *Criminal Justice*, Yongin University, South Korea Thesis: *The effectiveness of integrating CCTV in South Korea*
2004-2008	B.S., *Public Administration*, Korean National Police University

PUBLICATION

Peer-Reviewed Journal

Oh. G, Song, J., Park, H., & Na, J. (2021). Evaluation of Random Forest in Crime Prediction: Comparing Three-Layered Random Forest and Logistic Regression, *Deviant Behavior,* Online Publication, 1-17.

Oh. G, Zhang, Y., & Greenleaf, R. (2021). Measuring Geographic Sentiment toward Police Using Social Media Data, *American Journal of Criminal Justice,* Online Publication, 1-17.https://doi.org/10.1007/s12103-021-09614-z

Na, C., Song, J., **Oh, G., &** Park, H. (2021). Do Machine Learning (ML) Methods Outperform Traditional Statistical Models in Crime Prediction?: A Comparison between Logistic Regression (LR) and Neural Networks (NNs), *Korean Journal of Policy Studies*.

Oh, G, Ren, L., & He, P. (2019). Social Disorder and Residence-based Fear of Crime: The Differential Mediating Effects of Police Effectiveness. *Journal of Criminal Justice,* 63, 1-11. https://doi.org/10.1016/j.jcrimjus.2019.05.001

Oh, G., & Connolly, E.J. (2019). Anger as a mediator between peer victimization and deviant behavior in South Korea: A cross-cultural application of general strain theory. *Crime and Delinquency*, 65(8), 1102-1122. https://doi.org/10.1177/0011128718806699

Kim, J**., Oh, G**., & Siennick, E. (2018). Unraveling the effect of cell phone reliance on adolescent self-control. *Children and Youth Services Review, 87*, 78-85.

Ha, T., **Oh, G**., & Park, H. H. (2015). Comparative analysis of defensible space in CPTED housing and non-CPTED housing. *International Journal of Law, Crime, and Justice, 43(4)*, 496-511.

Oh, G., Kim, T., & Park, H. H. (2013). A study on effectiveness of integrated CCTV control system: With focus on monitoring operators and investigators. *Journal of Korea Science and Art Forum, 13,* 219-232. (In Korean) http://kiss.kstudy.com/thesis/thesis-view.asp?key=3207984

Park, H. H., **Oh, G.**, & Paek, S. Y. (2012). Measuring the crime displacement and diffusion of benefit effects of open-street CCTV in South Korea. *International Journal of Law, Crime, and Justice, 40(3)*, 179-191.

Park, H. H., & **Oh, G.** (2010). Study on criminal profiling system in South Korea: The perception of new police profilers and career detectives. *Korean Police Studies Review, 9(2),* 59-88. (In Korean) https://www.earticle.net/Article/A126781

Under Review

Oh, G., & Connolly, E. J. (*Under review*) The role of Anger between Perceived Neighborhood Disadvantage, Offending, and Contact with the Criminal Justice System

Song, J., **Oh, G., &** Song, T. (*Under review*) Predicting Juvenile Recidivism using Machine Learning: Focusing on Korean Juvenile Offenders

In Progress

Oh, G., Z, Y., & R, Greenleaf., Measuring Neighborhood Physical Disorder using Non-Emergency Calls

ACADEMIC AND PROFESSIONAL POSITIONS

2020-Present	Researcher National Police Science Institute, Korean National Police University
2019-2020	Doctoral Teaching Fellow Department of Criminal Justice and Criminology, Sam Houston State University
2018-2019	Graduate Research Assistant Department of Criminal Justice and Criminology, Sam Houston State University
2017-Present	Graduate Teaching Assistant Department of Criminal Justice and Criminology, Sam Houston State University
2016-2017	Detective (Lieutenant) Sexual Violent Unit of Gwangmyung Police Station, South Korea
2015-2016	Teaching Assistant

	Department of Criminology and Criminal Justice, Florida State University
2014-2015	Distance Research Assistant Korean National Police Agency, Seoul, South Korea *Research on improvement of dispatching police patrol officers*
2013-2014	Distance Research Assistant Korean National Police University, Seoul, South Korea *Research on data mining method of crime statistics in South Korea*
2011-2013	Detective (Lieutenant) Serious Crime Unit of Gyeonggi Police Agency, South Korea
2010-2011	Police Management Officer (Project Manager) Gwangmyung Police Station, South Korea *Integrating automobile tracking systems and public surveillance cameras*
2008-2010	Police Officer (Deputy Commander) Riot Police Unit of Seoul Metropolitan Police Agency, South Korea

RESEARCH INTERESTS

Computation Criminology using Big Data and Machine Learning

Biopsychosocial Criminology, Developmental Criminology

Policing

PROFESSIONAL CONFERENCE PARTICIPATION

Paper Presentations

2019	**Oh, G**., Zhang, Y. Assessing the Measurement of Collective Attitude Toward Police Using Big Data, American Society of Criminology, San Francisco, CA.
2019	**Oh, G**., & Song, J. Delinquency Forecast using Google Search Trends. Asian Criminological Society, Cebu, Phillipine.
2019	**Oh, G,** Ren, L., & He, P. Social Disorder and Residence-based Fear of Crime: The Differential Mediating Effects of Police Effectiveness, Academy of Criminal Justice Science, Baltimore, MA.

2019	**Oh, G**., Zhang, Y. Crime Prediction Using Administrative Big Data and Machine Learning Algorithms, International Conference on Big Data and Data Science, Barcelona, Spain.
2018	**Oh, G**., & Connolly, E. J. The role of anger between perceived neighborhood disadvantage, offending, and contact with the criminal justice system. American Society of Criminology, Atlanta, GA.
2017	**Oh, G.** The effects of bullying victimization on substance use and deviant behaviors. American Society of Criminology, Philadelphia, PA.
2016	**Oh, G.,** & Park, I. Gentrification and crime rates. American Society of Criminology, New Orleans, LA.
2015	Stults, B. J., **Oh, G.**, & Swagar, N. The conditional effects of neighborhood context and parental effectiveness on self-control. American Society of Criminology, Washington D. C.
2015	Kim, J., & **Oh, G.** Cellphone reliance: Low self-control and delinquency. American Society of Criminology, Washington D. C.
2015	**Oh, G.** The effectiveness of integrated surveillance system. Academy of Criminal Justice Science, Orlando, FL.
2014	**Oh, G.** Is intelligent surveillance system effective? American Society of Criminology, San Francisco, CA.

HONORS, AWARDS, AND SCHOLARSHIPS

2019	SHSU, Raven Scholar Award
2019	SHSU, Graduate Research Summer Fellowship ($6,672)
2019	SHSU, Dan Beto Correctional Leader Scholarship ($2,000)
2019	SHSU, Graduate Studies Scholarship ($1,000)
2018	SHSU, Graduate Research Summer Fellowship ($6,672)
2018	SHSU, Dan Beto Correctional Leader Scholarship ($2,000)
2017	SHSU, Graduate Studies Scholarship ($1,000)
2013	Florida State University, Dean's Scholarship ($1,000)

2012	Yongin University, South Korea, Honor Graduate Student Award
2009	Chief of the Mobile Police Regiments, Commendation for Leadership
2007	Korea National Police University, South Korea, Outstanding Undergraduate Student Award

SERVICE

Professional

Manuscript Reviewer

- *Journal of Research in Crime and Delinquency*
- *Journal of School Violence*
- *Criminology and Public Policy*
- *International Justice of Law and Criminal Justice*

University

2018	SHSU, Peer Mentorship: M.A. Student Abigail Eck
2018	SHSU, Statistics Tutoring: First year Ph.D. Student Cohort
2015-2016	Florida State University, Secretary, Graduate Student Organization

GRANT EXPERIENCE

2018 Internal Grant, SHSU
Measuring collective sentiment and social disorder using Twitter
Directed by Dr. Yan Zhang

TEACHING EXPERIENCE

Instructor

2020	Sam Houston State University- Introduction to methods of research (Online)
2019	Sam Houston State University- Introduction to methods of research

Guest Lectures

2018	Sam Houston State University- Qualitative and quantitative methods
2018	Sam Houston State University- Policy evaluation research
2016	Florida State University- Ethics in social science research
2016	Florida State University- Validity, reliability, and scale development
2013	Yongin University- Police investigation

PROFESSIONAL DEVELOPMENT

Computer Programming Proficiencies

- R
- Python

Statistical Program Proficiencies

- STATA
- MPLUS
- SPSS
- AMOS
- Esri Arc GIS
- SAS
- HLM

PROFESSIONAL AFFILIATIONS

- American Society of Criminology
- Academy of Criminal Justice Sciences
- Division of Biopsychosocial Criminology, American Society of Criminology