Ethicality of Data Mining and Predictive Analytics

SOPHIE ROPER, PARK UNIVERSITY
Mentor: Dr. Adrian James, Department Chair
and Assistant Professor of Management, Park University

Introduction to Practices

Big data plays an increasingly important role in our daily lives. There are four dimensions of big data: volume, velocity, variety, and veracity. Due to continuous technological advancements, data collection and analysis are exponentially increasing (Executive Office of the President, 2014). According to predictive analyst Seigel, data increases by nearly 2.5 quintillion bytes a day (2013). President Obama's 2014 Data Privacy Report notes that volume, velocity, and variety are all increasing (Executive Office of the President, 2014). There are concerns that have arisen with the prevalence of big data in our lives. Two of the primary big data practices utilized are data mining and predictive analytics.

According to founders of Data Mining Inc., Linoff and Berry, data mining is "the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules" (Linoff & Berry, 2011, p. 3). Additionally, there are two primary methods of data mining, directed and undirected. Directed strives to explain or categorize a particular field, while undirected endeavors to find patterns among groups of data without the use of a particular field. The following tasks can all be performed through data mining: classification, estimation, prediction, affinity grouping, clustering, and profiling (Linoff & Berry, 2011). Data mining works under the assumption that data from the past will be relevant to the future. While "data mining" is frequently used synonymously with predictive analytics, it is more of "digging around" through data (Siegel, 2013, p. 12).

Predictive Analytics is "technology that learns from experience (data) to predict the future behavior of individuals in order to drive better decisions" (Siegel, 2013, p. 11). In essence, data is fed to a machine capable of learning and that machine then determines predictions about the future (Siegel, 2013). Predictive analytics can tell us a number of things, including Rhianna listeners tend to vote Democrat, those who retire sooner have shorter life expectancies, or even those who eat meatless diets miss less flights (Siegel, 2013). Industries utilize these learning machines as well. For instance, prediction can be used to strengthen healthcare, combat crime, and more accurately target ads (Siegel, 2013). While predictive analytics seemingly predicts the future, prediction based on past data is implemented in most decision-making systems (Ferguson, 2017).

Specific fields in which these practices are employed include healthcare, law enforcement, and marketing. A primary question that arises with the use of data mining and predictive analytics within these fields is whether or not they are ethical. The practices have shown results in many areas, but there are also numerous suspected consequences. It must be determined whether or not the results outweigh the harms done.

Marketing

It is common practice for retailers to accumulate data on consumers without their knowledge. Many large retailers study purchases to directly target advertisements to consumers. Big data is used in retailing along the following dimensions: “data pertaining to customers, products, time, (geo-spatial) location and channel” (Bradlow et al., 2017).

Hypothesis 1: Using data mining and predictive analytics for marketing purposes provides convenience at cost of individual privacy.

One such expert at data mining is Target (Hill, 2012). Led by their expert statistician, Andrew Pole, Target assigns each customer an ID number. Attached to this ID number is personal information such as credit card number, email address, and phone number. The ID number also monitors purchase history. Using these ID numbers, Pole cross referenced baby registers to ascertain what products pregnant women were buying and when. With 25 popular products appearing consistently within the registers, Target was then able to assign shoppers without a registry a pregnancy score (Hill, 2012). The predictive analytics application for the situation is as follows: 1.) What is predicted? Which customers are likely pregnant 2.) What does Target do? Target sends targeted advertisements to those likely to be expectant parents (Siegel, 2013).

If they bought enough of the popular items in the baby registries, Target concluded that they were likely pregnant. After accumulating so much data on individuals, Target begins to target their ads. Those with high pregnancy scores began receiving advertisements for baby products (Hill, 2012). This demonstrates how data can be compiled to determine an individual’s preferences. Such profiles can help segment consumers into set groups, which will receive specialized advertisements (Executive Office of the President, 2014).

These highly targeted advertisements do not always go over well. One man became enraged that his teenage daughter was receiving advertisements from Target for pregnancy related items such as cribs and baby clothes. After contacting customer service, they apologized numerous times. Upon hearing the last apology, the man changed his tone. His daughter was, in fact, pregnant (Hill, 2012). This means that the high pregnancy score assigned to his daughter was accurate. Target knew his teenage daughter was pregnant before he did. Upon realizing Target knowing about a pregnancy before the immediate family might alarm some people, they changed tactics. They now mix in non-pregnancy related items into the targeted ads so that the products advertised appear to be shown randomly (Hill, 2012). The pregnancy items are still heavily advertised, but to the undiscerning eye, it is much more subtle.

Prediction is also utilized on the coupons advertised on receipts. In the UK, grocer Tesco attempts to determine which discounts will be most utilized to personalize coupons (Siegel, 2013). This practice has been shown to have success. The redemption rate for coupons has nearly quadrupled using this practice. Many other grocery chains have followed Tesco's lead (Siegel, 2013). Predictive analyst Siegel detailed his own embarrassing experience when his Walgreens receipt provided him with targeted coupons for Beano. They were able to determine this as a potential product of interest after he purchased products relating to his lactose intolerance (Siegel, 2013).

For online targeted advertising, what is being predicted is which ads customers are most likely to engage with (Siegel, 2013). The likelihood of a user engaging with an advertisement, i.e. clicking on it, depends on a number of factors, including "the individual's current school year, gender, and e-mail domain" and more (Siegel, 2013, p. 25). Then, based on this information, the sites display the best ad suited for your preferences (Siegel, 2013). This targeting method is done gradually, often without user notice. These targeted ads have been shown to boost revenue (Siegel, 2013).

Such personalization is not only open to privacy violations, but to discrimination. One study found that internet searches with "non-white sounding" names more frequently displayed advertisements with "arrest" than searches with "white sounding" names (Executive Office of the President, 2014, p. 7). Racist targeted advertising is, unfortunately, not a new concept. The tobacco industry was notorious for targeting minorities. These marketing tactics were then passed on to the food and beverage industry. They target at-risk populations, especially those with obesity disparities (Nguyen, Glantz, Palmer, & Schmidt, 2020). In 2017 alone, over one billion dollars was spent on advertisements using Spanish language. 80% of these ads promoted unhealthy foods (Harris, 2020).

Stores also know which items to stock at certain times of the year based on information they have gathered. After performing an analysis, Walmart learned that when poor weather is forecasted, three items face a major rise in demand: duct tape, bottled water, and strawberry pop-tarts (Pearsall, 2010). These items are now more heavily stock in preparation for poor weather conditions.

While there are several drawbacks as discussed, there are an abundance of benefits linked to the increase in targeted advertising. Targeted advertising has boosted the software and consumer application industry immensely. These practices allow marketers to better engage with their customers in ways they want to be reached (Parkin & Hoopla, 2014).

Conclusion 1: Predictive analytics and data mining in marketing are worthwhile practices. The benefits of these practices outweigh the invasion of privacy and occasional instances of racial profiling. Ultimately, the practices allow marketers to better target ads so that customers see products of most interest.

Law Enforcement

Predictive tools were meant to make risk-based assessment decisions easier (Ferguson, 2017). The intent of predictive policing, mining data and using predictive analytics to detect crime, is to prevent the crime before it even happens. Los Angeles, Santa Cruz, and Iraq were among the first to implement modern predictive policing. In Iraq, studies showed that predictive policing predicted crime 262 percent more correctly than hotspotting. These results are unable to be fully valued, as they were conducted by PredPol, a top predictive policing firm (Bond-Graham, 2013). Santa Cruz police department used data mining to gather data on burglaries and then applied a predictive algorithm to determine dates and locations of future robberies. Police then patrol said areas during the predicted time for the burglary to occur. Social maps link known criminals together and target them heavily (Baxter, 2012). Predictive policing was meant to be the future of law enforcement, removing human biases from the equation.

Hypothesis 2: Using data mining and predictive analytics in law enforcement is an effective measure, but can be reliant on racism and stereotyping, perpetuating racial injustice.

To some extent, prediction is inherent in policing. It is not uncommon for detectives to follow a hunch based on precedent. The major change then, is in the resources utilized to predict crime, not the strategy of prediction itself (Ferguson, 2017). Crime maps have matured from push pins on a map to digitalized histories of crime in certain areas. The general strategy is to deploy more officers to areas on crime maps with higher likelihood of crime occurring, especially crimes like theft and battery (Ferguson, 2017).

Early cases of utilizing prediction in criminal justice were for parole recidivism in Chicago during the 1920s. Ernest Burgess and his fellow adopters looked at risk factors to predict whether or not parolees would perpetrate another crime. Burgess's practices became known as Actuarial Prediction and began to replace the "Clinical Method" (Burgess, 1928). In contrast to the Actuarial Method, the

Clinical Method did not consider pre-identified variables. Current examples of actuarial prediction used within law enforcement include the Violence Risk Appraisal Guide, measuring recidivism rates for those with mental illness (Ferguson, 2017).

There have been three general iterations of predictive policing. The first, predictive policing 1.0, "involved the collection of historical crime data and the application of an experimental computer algorithm that used data to predict likely areas of criminal activity" (Ferguson, 2017). This iteration relies upon the ripple effect, the idea that property-based crimes will spread to neighboring areas. Environmental Criminologists Frith, Johnson, and Fry explain that the physical environment is crucial to crime patterns. Criminals' decisions are heavily influenced by the modern road system as locations that are near their base of operations, are familiar, and are prone to less witnesses are primary locations identified as being at risk of burglary (Frith, Johnson, & Fry, 2017). Questions regarding the validity of studies conducted on predictive policing 1.0's success make it impossible to gauge its true success (Ferguson, 2017). Perceived success of the practices provided a springboard for future iterations of predictive policing.

The second iteration of predictive policing, predictive policing 2.0, moved away from preventing property crimes such as burglary towards preventing violent crimes such as shootings and robberies (Ferguson, 2017). They both, however, focus on locations of crime. With funding from the Bureau of Justice Assistance, numerous pilot programs were launched in cities like Baltimore, Las Vegas, Los Angeles, and Kansas City. In Boston, for instance it was discovered that less than five percent of street corners accounted for nearly three fourths of shootings between 1980 and 2008. Other factors like time of day (after school is out, weekend evenings, etc.) were also deduced (Braga et al, 2010). In Philadelphia, police began old-fashioned foot patrols of areas that were predicted to have the highest crime. With the help of Dr. Jerry Ratcliffe, the Philadelphia Foot Patrol Experiment began in 2007. Ratcliffe found that after three months of officers walking their beats, violent crime decreased by 23% compared to areas without foot patrols (Ratcliffe & Sorg 2017).

Since the mid 1990s, New York City's crime rate has steadily declined (Gladwell, 2009). Crime causes such as gentrification and drug trade have tapered off, so there must be another reason for the decrease in crime. One possibility is the computerized map the NYPD implemented under Commissioner Raymond Kelly, that is used to show where serious crimes are occurring in real time. Officers are then sent to these "impact zones" at increased rates. With this practice, these zones averaged a 35% decrease in crime (Gladwell, 2009, p. 632-633). There have also been increases in racial profiling and police targeting of lessor crimes. As David Robinson, Upturn think tank creator, says, "People in heavily policed communities have a tendency to get in trouble. These systems are apt to continue those patterns by relying on that biased data" (Jouvenal, 2016). Those most targeted and reported are not necessarily the ones committing the majority of crimes.

Predictive Policing 3.0 shifted from targeting locations where crime is likely to occur toward people likely to perpetrate crime. Factors include "past criminal activity, current associations, and other factors that correlate with criminal propensity" (Ferguson, 2017). This iteration is engrained in the discovery that a small percentage of the population carries the largest risk of being involved in gun violence, whether as the victim or the perpetrator. In Chicago, for instance, most victims and offenders "tend to be young African American men who live in neighborhoods on the West or South sides of the city" (Meares et al, 2009). This "Heat List" of people likely to commit a crime was shown to be accurate when on Memorial Day weekend of 2016, 78% of the 64 individuals shot were identified (Papachristos, 2016). Yet, overall violence in the city has not decreased, as the list of individuals are treated as suspects, and little effort is taken to implement violence prevention programs (Ferguson, 2017). A 2013 RAND Corporation study even found that the individuals on the lists were no more likely to be tied to shootings than any other group. Several police departments dispute RAND's findings, however (Jouvenal, 2016).

In Kansas City, Missouri, a social network analysis was performed for the most probable offenders. The initial target list included murder, shooting, and assault suspects. This list was then analyzed to identify their associates. The new list of potential offenders and their respective associates was then examined again, for associates of the associates. This was performed once more so three layers of offenders were identified. The final list of 120 individuals were contacted by police and made aware of social programs to lower their chances of offending. When individuals from the list of 120 did offend, they were punished more harshly than those not on the list (Braga et al, 2010). Similar practices have been utilized in other cities. In Los Angeles, California, Operation LASER identifies likely offenders and forms a bulletin similar to a most wanted list. The bulletins are then distributed to officers for surveillance and investigation. New Orleans partnered with data software firm Palantir to analyze gang activity, drugs, feuds, and other factors to determine those most likely to be involved in gun violence. They identified about 3,000 individuals. Upon acting on these tips, the murder rate dropped 21.9% (Ferguson, 2017). This correlation between the use of predictive policing software and decreased crime does not necessarily mean that predictive policing is the cause (Perkowitz, 2016). Another drawback is that while attempting to branch into violent crime, policing softwares currently focuses primarily on theft.

It's important to note that while these technological "improvements" provide officers with individuals to surveil and investigate, they themselves do not provide just cause to stop or arrest (Ferguson, 2017). While many areas report lower crime rates using predictive policing, many ethical questions have arisen, namely whether or not individuals are being falsely accused of crimes based solely on who they affiliate themselves with.

It must be determined whether or not software that police have acknowledged targets individuals more than crimes can be unbiased. If not, the software could be a civil liberties violation (Bond-Graham, 2013). This is by no means the first time police departments have been accused of misconduct. Between 1986 and 1990, 1800 of the 8500 officers in the LAPD had allegations of excessive force or improper conduct made against them (Gladwell, 2009). The predictive policing hopes to help officers strengthen community relationships which have become particularly tarnished in recent years (Perkowitz, 2016). Many fear that predictive policing software will promote distrust of police. Many activist groups, namely the ACLU, predict the software will unfairly target vulnerable minorities and threaten fourth amendment protections (Perkowitz, 2016).

The primary analysis conducted on predictive policing has been completed by the predictive policing firms. The limited research done by third parties proves to be inconclusive. For instance, Brantingham and colleagues claim Predpol is twice as good at predicting where crime will occur than crime analysts from the LAPD, but no outside researcher can confirm (Jouvenal, 2016). HunchLab, a competitor predictive policing firm, has made similar claims. Part of the reason for lack of third-party research is the practices evolving at a quicker rate than debates regarding the practices. The public is rarely even made aware the police departments in their cities are partnering with predictive companies. Even when the practices are analyzed, the three iterations are typically analyzed together. Whether shown to be effective or not, there has been an increasing trend of departments adopting predictive policing methodologies. Part of the seemingly blind support for predictive policing firms is contractual obligations of police departments to endorse the software to colleagues and speak about it at press conferences (Bond-Graham, 2013).

A major factor contributing to potential inaccuracies and injustices within predictive policing is flawed data (Ferguson, 2017). Human error is a major contributing factor. This could be something as simple as an officer taking down a name wrong. Some police departments plainly do not have enough data to predict crime patterns. With smaller sample sizes, predictive judgements become less accurate. National numbers cannot be accurately utilized because they are not representative of any given city. Biased data is also a primary concern. Victims are more likely to report crimes such as auto theft and burglary, but less likely to report domestic violence and sexual assault. Racism, unsurprisingly, can also cause bias. "The targeting of certain areas or certain races creates the impression of higher crime rates in those areas, which then justifies continued police presence there" (Ferguson, 2017). This is not to say that all police let their biases influence the data, but as was seen from 1986 to 1990, those at one extreme on the spectrum make a large impact on the total population (Gladwell, 2009). Certain communities do report higher crime, but this could well be because they are watched more (Ferguson, 2017).

Biases are often not even rooted in truth. Just as it is assumed crime will be perpetrated by people with lower income, in predominantly black neighborhoods, it is assumed that those associated with the mafia are Italian. A small fraction of Italian Americans are members of the mob (Gladwell, 2009). This is in the same way that there is no hard evidence that those who are African American and/or live in lower income neighborhoods are more likely to commit crime.

In many ways, the program then tells you what you want to see. If you only input information about minority areas, they will appear to be problem areas. Ultimately, the data collected and imputed into the system is based upon human judgement (Ferguson, 2017). It can be argued, however, that predictive policing data is more accurate than arrest records, as it considers victim reports directly. Further problems can occur if the crime is reported into the system twice, once from arrest records, and once from a victim report (Ferguson, 2017).

There are ways to amend the data collection process. The first step is to acknowledge bias and error within the data (Ferguson, 2017). As within any law enforcement profile, it is merely a starting basis, not an exact guidebook to follow. Vulnerabilities within data can be caught and corrected through auditing mechanisms. This can be difficult due to the extensive amount of data being collected, but practices such as finding and deleting duplicate data can go a long way (Ferguson, 2017). Additionally, most officers are not trained in data management, so training and technology must be utilized to assist them. Departments hiring crime analysts can also help to alleviate the burden. Also important is making sure systems are up-to-date and working as efficiently as possible. One question regarding predictive policing is why not focus on police activity as well (Ferguson, 2017). For instance, where were they when crimes occurred. This police-centric approach could go a long way to remove potential biases, creating a more complete picture of crime (Ferguson, 2017).

It seems as if for every successful predictive policing story, there are several failures. It is important to note that what works in one city, may not work for another. This is due to "differences in geography, crime patterns, or police culture" (Ferguson, 2017). Even Jeffrey Brantingham, a founder of modern policing and creator of predictive firm PredPol stated when speaking to Next City contributor Christopher Moraff, "'Person-centric' models are problematic...because they carry an elevated margin of error and can legitimize racial, gender-based and socioeconomic-driven profiling" (Moraff, 2014). He goes on to state that the model can result in many false positives, noting someone as a suspect when they are innocent

(Ferguson, 2017). The full validity of predictive policing may never be known. Even though PredPol, HunchLab, and other firms are beginning to analyze their programs, criminology is an imperfect testing ground (Ferguson, 2017). There are too many constantly changing variables. The few independent assessments of predictive policing technology that have been done are inconclusive. Criminologist Ed Schmidt, a former police officer, says that while he believes in the concept of predictive policing, he has qualms about its current claimed effectiveness (Bond-Graham, 2013).

Conclusion 2: Predictive policing and data mining are laudable practices when conducted ethically. Discriminatory practices are primarily due to human bias, shortcomings of laws, and the algorithms allowing the introduction of human bias. To make practices more efficient, auditing and data protection needs to be implemented. This is especially true in law enforcement, where the stakes for usage are much higher. Some police do use predictive policing to target minorities. This is an obvious example of an unethical practice. This is an exception, not a standard, however. Overall, the practices do help catch criminals.

Healthcare

Predictive medicine is an emerging big data field. Healthcare's data driven nature has been fostered by legislation, which, in past years, have provided incentives for the transition to electronic records. This conversion has progressively improved the volume of data available to healthcare professionals (Executive Office of the President, 2014). With predictive medicine, doctors are able to better predict which individuals are most at risk for certain diseases and what treatments would best remedy them. This is accomplished through analysis of existing health data and genetic information. The benefits of this could be great for those with similar genetics and the patient's descendants (Executive Office of the President, 2014). Having access to more genetic data was found by the Broad Institute to make crucial differences in identifying variations in genes indicating disease. According to "Big Data: Seizing Opportunities, Preserving Values" from The Executive Office of the President, a "genetic variant related to schizophrenia was not detectable when analyzed in 3,500 cases, and was only weakly identifiable using 10,000 cases, but was suddenly statistically significant with 35,000 cases" (Kellis, 2014). This illustrates the potential life saving capabilities of having access to more data. It is not just the collection of data that matters. There are many insights to be gleaned to improve healthcare outcomes (Kumar, 2018).

Hypothesis 3: The use of predictive analytics and data mining in healthcare eradicates discriminatory practices

Obtainment and analysis of data have been essential to improving society. One early example was in the 1800s, when Dr. John Snow mapped "clusters" of cholera within London. These clusters helped to identify the source of the cholera outbreak (Tulchinsky, 2018). A more recent example is data collected from neonatal intensive care unit monitors helping to determine which newborns are most likely to contract serious infections. The system is able to analyze more data than doctors can gather on their rounds, such as constant temperature and heart rate. Spikes in these vitals can be indicative of infection occurring in the body (IBM, 2012).

There are also big data practices that can be utilized to prevent patients from needing to visit the doctor, going beyond the apple a day adage. Healthcare professionals are able to identify lifestyle factors such as diet and exercise that are likely to keep patients healthier. By following such practices, the predicted need for doctor visits greatly decreased. For times when a doctor's appointment is necessary, proper treatments and prescriptions are more identifiable (Executive Office of the President, 2014). When admittance to the hospital is required, it can be difficult to triage. Predictive algorithms could run risk analysis on patients to decide who most needs ICU care (Cohen, Amarasingham, Shah, & Lo, 2014).

There are some practices already in place to prevent mistreatment of patient data. HIPAA, the Health Insurance Portability and Accountability Act established in 1996, states that only necessary individuals can use and share a patient's health information. This is intended to help patients understand and control how their data is used (Executive Office of the President, 2014). It is questioned if HIPAA is enough, however. Individuals' health data is frequently shared, and even sold (Executive Office of the President, 2014). Another safeguard is model developers including patients and stakeholders in governance structures from the beginning of development (Cohen, Amarasingham, Shah, & Lo, 2014).

Analytics practices address many pressing issues within hospitals, such as patient safety, patient communication, and patient information (Miner et al, 2015). An example of analyzing patient information is the Centers for Medicare and Medicaid Services utilizing predictive analytics to determine who is likely to commit reimbursement fraud. $115 million in fraudulent payments have been prevented and identified (Executive Office of the President, 2014). Healthcare professionals are able to form conclusions that would not have been previously possible thanks to digitized and new data. These conclusions can range from medical diagnosis to who is likely to pay their bill.

Conclusion 3: Data mining and predictive analytics are useful practices within healthcare with few HIPAA concerns. The practices help healthcare workers make better decisions when it comes to diagnosis and triage

Recommendations and Conclusions of Practices

It is crucial to note that neither data mining or predictive analytics are inherently bad. As Melvin Kranzberg's First Law of Technology states, "Technology is neither good nor bad; nor is it neutral" (Kranzberg, 1995). There are already some measures in place to ensure the practices of data mining and predictive analytics are used ethically. Fair Information Practice Principles (FIPP), for example, established basic personal data protections. They state that an individual has the right to know what information is being collected and how that information will be utilized. The principles go on to state that individuals should be able to protest certain data exploitations. FIPP set precedent for the Privacy Act of 1974, which "prohibits the disclosure of a record about an individual from a system of records absent the written consent of the individual, unless the disclosure is pursuant to one of twelve statutory exceptions" (*Privacy Act 1974*). Healthcare data is often subject to conflicting federal and state regulation though. Many argue that this makes it necessary for more specific legislation regarding healthcare data to establish universal standards (Executive Office of the President, 2014). This is a legitimate need since individual hospitals are becoming coordinated data systems (Burroughs, 2019).

While some fear an invasion of privacy with data mining and analytics, it is ultimately just drawing conclusions based on the facts already collected (Siegel, 2013). Siegel argues if he were to "glance into my friend's shopping cart and, based on certain items, draw the conclusion that she may be pregnant, have I just committed a thoughtcrime—the very act enforced against by Big Brother in George Orwell's Nineteen Eighty-Four?" (Siegel, 2013, p. 65). The argument can be made that predictive analytics do not differ much from this scenario. Predictive analytics merely takes advantage of all data present.

It has been suggested by many experts that legislation on data protection and discrimination is necessary (Favaretto, 2019). This is largely due to the fact that racial biases are cyclic, and predictive policing serves to continue the cycle (Siegel, 2013). In theory, data mining should be more objective than human decision-making as it focuses on empirical data that ought to promote objectivity. Objectivity could potentially "limit human error and bias" (Favaretto, 2019). Even relying on empirical data, data mining is not without human interaction, as the results must be reviewed for accuracy. Some experts go so far as saying data mining could be utilized to discover practices of discrimination used within the practice itself (Favaretto, 2019).

Another solution would be removing certain data that makes individuals identifiable to the human user or even letting individuals dictate how their data is used. While these may limit discrimination, they could also remove usefullness from the data (Executive Office of the President, 2014). Ellen Kurtz, a proponent for predictive policing, stated to Nadya Labi for an *Atlantic* article, "If you wanted to remove everything correlated with race, you couldn't use anything. That's the reality of life in America." (Labi, 2012). Discrimination is a concern in healthcare as well, as those who are already disadvantaged are at risk of becoming even more so (Cohen, Amarasingham, Shah, & Lo, 2014).

Predictive analytics can be bettered by the Ensemble Effect. This occurs when several predictive models are used to compensate for any one model's limitations. Such compensation results in more accurate predictions (Siegel, 2013). Still, it has been stated that absolute accuracy is unachievable. This begs the question whether or not prediction is worth it. As it turns out, predictions do not have to be completely accurate to have value (Siegel, 2013). This is best demonstrated in advertising. If an algorithm sends targeted ads for protein powder to a small number of people who do not use protein powder, little is lost. Some of those people may even see the sale as an initiative to try the product for the first time. Similar benefits occur with advertisements sent via mail. By targeting only certain individuals to send ads to, the company saves money by not sending advertisements to those not likely to respond well. Also, the likely non-responders are spared the additional junk mail (Siegel, 2013). These scenarios were examples of the prediction effect. The prediction effect states that any prediction, even if not completely accurate, is better than guessing and has notable value (Siegel, 2013).

Implementations for Practice

This discussion demonstrates that practices can ethically be used in each of the three aforementioned fields, as long as fail-safes are practiced. Those running the programs will determine their ethicality. If reviews, upgrades, and ethical use practices are utilized, data mining and predictive analytics are not only effective, but ethical. The use of data mining and predictive analytics could be improved by making consumers more cognizant about the collection of their data. If there were opt-in and opt-out options, consumers would have more control over their data. For law enforcement, bias training would correct most of the current issues seen within predictive policing. Even with bias training, however, the biased nature of the programs themselves would still need to be addressed. Additionally, more research needs to be conducted to study the impacts of predictive policing. It is indisputable that predictive policing increases arrests, but less certain whether or not it actually decreases crime. Within healthcare, following HIPAA guidelines should be the primary concern. As long as these guidelines and medical triage are followed, providing those with the most urgent needs with priority care, little change needs to be made to current practices.

References

Baxter, S. (2012) Modest Gains in First Six Months of Santa Cruz's Predictive Police Program. Santa Cruz Sentinel.

Bond-Graham, D. (2013, October 30). All Tomorrow's Crimes: The Future of Policing Looks a Lot Like Good Branding. SF Weekly. https://www.sfweekly.com/news/all-tomorrows-crimes-the-future-of-policing-looks-a-lot-like-good-branding/.

Bradlow, E., Gangwar, M., Kopallec, P., & Voletib, S. (2017). The Role of Big Data and Predictive Analytics in Retailing. Journal of Retailing, 93(1), 79–95. https://doi.org/https://doi.org/10.1016/j.jretai.2016.12.004

Braga, A. et al. 2010. The Concentration and Stability of Gun Violence at Micro Places in Boston, 1980–2008, 1 Journal of Quantitative Criminology 26, 33–53.

Burgess, E. W. 1928. "Factors determining success or failure on parole." In Bruce, A. W., E. W. Burgess, J. Landesco, & A.J. Harno (Eds.), The Workings of the Indeterminate Sentence Law and the Parole System in Illinois.Springfield: Illinois Board of Parole, pp. 221-234

Burroughs, J. (2019). Essential Operational Components for High-Performing Healthcare Enterprises. Health Administration Press.

Cohen, I. G., Amarasingham, R., Shah, A., Xie, B., & Lo, B. (2014). The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care. Health Affairs, 33(7). https://doi.org/https://doi.org/10.1377/hlthaff.2014.0048

Executive Office of the President. (2014). Big Data: Seizing Opportunities, Preserving Values.

Favaretto, M., De Clercq, E., & Elger, B. S. (2019). Big Data and discrimination: perils, promises and solutions. A systematic review. J Big Data 6, 12.

Ferguson, A. (2017). Policing Predictive Policing. Washington University Law Review, Volume 94, 1109.

Frith, M. J., Johnson, S. D., & Fry, H. M. (2017). Role of the Street Network in Burglars' Spatial Decision-Making. Criminology, 55(2), 344-376. https://doi.org/10.1111/1745-9125.12133

Gladwell, M. (2009). What the Dog Saw: and Other Adventures. Little, Brown and Company.

Harris, J. L. (2020). Targeted Food Marketing to Black and Hispanic Consumers: The Tobacco Playbook. American Journal of Public Health, 110(3), 271–272. https://doi.org/10.2105/AJPH.2019.305518=

Hill, K. (2012). How Target figured out a teen girl was pregnant before her father did. Forbes, Inc.

IBM. (2012). Smarter Healthcare in Canada: Redefining Value and Success. http://www.ibm.com/smarterplanet/global/files/ca__en_us__healthcare__ca_brochure.pdf

Jouvenal, J. (2016, November 17). Police are using software to predict crime. Is it a 'holy grail' or biased against minorities? The Washington Post. https://www.washingtonpost.com/local/public-safety/police-are-using-software-to-predict-crime-is-it-a-holy-grail-or-biased-against-minorities/2016/11/17/525a6649-0472-440a-aae1-b283aa8e5de8_story.html.

Kellis, M. (2014). Importance of Access to Large Populations. Big Data Privacy Workshop: Advancing the State of the Art in Technology and Practice, Cambridge, MA. http://web.mit.edu/big-data- priv/ppt/ManolisKellis_PrivacyBigData_CSAIL-WH.pptx.

Kranzberg, M. (1995). Technology and History: "Kranzberg's Laws." Bulletin of Science, Technology & Society, 15(1), 5–13.

Kumar, V. (2018). Healthcare Analytics Made Simple : Techniques in Healthcare Computing Using Machine Learning and Python. Packt Publishing.

Labi, N. (2012). Misfortune Teller The Atlantic. www.theatlantic .com/magazine/archive/2012/01/misfortune-teller/8846/.

Linoff, G. S. & Berry, M. J. A. (2011). Data mining techniques for marketing, sales, and customer relationship management. Wiley Pub., Inc.

Meares, T. et al. (2009). Project safe neighborhoods in chicago, homicide and gun

violence in chicago: evaluation and summary of the project safe neighborhoods program 2. http://www.saferfoundation.org/files/documents/2009-PSN-Research- Brief_v2.pdf.

Miner, L., Bolding, P., Hilbe, J., Goldstein, M., Hill, T., Nisbet, R., Walton, N., & Miner, G. (2015). Practical Predictive Analytics and Decisioning Systems for Medicine : Informatics Accuracy and Cost-Effectiveness for Healthcare Administration and Delivery Including Medical Research. Academic Press.

Moraff, C. (2014). The Problem with Some of the Most Powerful Numbers in Modern

Policing. Next City.

Nguyen, K. H., Glantz, S. A., Palmer, C. N., & Schmidt, L. A. (2020). Transferring Racial/Ethnic Marketing Strategies From Tobacco to Food Corporations: Philip Morris and Kraft General Foods. American Journal of Public Health, 110(3), 329–336. https://doi.org/10.2105/AJPH.2019.305482

Papachristos, A. (2016). Commentary: CPD's Crucial Choice: Treat Its List as Offenders or as Potential Victims?. Chicago Tribune.

Parkin, G. & Hoopla digital. (2014). Digital Marketing : Strategies for Online Success. IMM Lifestyle Books.

Pearsall, B. (2010). Predictive policing: The future of law enforcement. National Institute of Justice Journal, 266(1), 16-19.

Perkowitz, S. (2016, October 27). Should we trust predictive policing software to cut crime? Aeon. https://aeon.co/essays/should-we-trust-predictive-policing-software-to-cut-crime.

Privacy Act 1974 (5 U.S.C.) 552a (USA).

Ratcliffe, J. & Sorg, E. (2017). Foot Patrol : Rethinking the Cornerstone of Policing. Springer.

Siegel, E. (2013). Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die. John Wiley & Sons, Inc. October 21, 2020

Tulchinsky, T. H. (2018). John Snow, Cholera, the Broad Street Pump; Waterborne Diseases Then and Now. Case Studies in Public Health, 77–99. https://doi.org/10.1016/B978-0-12-804571-8.00017-2