Reflection of Experience Designing Mobile Application for Florida Police Departments

Analysis of Bias in Crime Prediction Algorithms

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Nikhil Sharma

October 27, 2022

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Benjamin Laugelli, Department of Engineering and Society

Mark Sherriff, Department of Computer Science

Introduction

In the past decade, mainstream media has been filled with an outrageous amount of news concerning police brutality, racial injustice, and other crimes. However, only a few weeks or months after these incidents happen media outlets taper their coverage of the topics and move on to a new source of news, causing no lasting change to be made to the institutions that seem to be disappointing us so often. The increase in crime in the United States over the past decade has led to many different software solutions that help police keep track of convicted felons, suspected criminals and other people in their jurisdiction. For the technical deliverable, a solution for police to keep track of crime in their jurisdiction was designed and implemented by the company I worked at: 22nd Century Technologies Inc. (TSCTI). TSCTI designed an app for police departments in 6 different counties of Florida that helped police officers keep track of ex-convicts and suspected criminals in their areas. Police officers are also able to submit new information to the app after they proof check this information with superiors to keep all other officers updated.

This increase in crime has also led to scrutiny for the police in the United States and questions about their training. (Mayson, S. G. , 2018). A modern day approach to the increase in crime is the use of algorithmic risk assessment. Algorithmic risk assessment uses machine learning to predict crimes using existing datasets that include decades of crime data. The issue with these algorithms is the data being used to train the algorithm as this data doesn't always reflect current day societal rules. Addressing the bias that exists in the data being used in these prediction algorithms can directly impact the amount of crime that is caught in the United States. This in turn would lead to the public supporting the police more as they can both identify crime and stop it earlier than they would be able to before. For the STS research, I will apply the theory

of technological politics to the case of Operation LASER, a LAPD algorithmic risk assessment model to show how the bias involved in these algorithms advantages certain groups while marginalizing others.

Technical

This technical research's primary focus is to reflect on my past internship experience . My internship at TSCTI helped me learn about the many software solutions that have been designed in recent times to help police keep track of criminal data and whereabouts, creating a pseudo-crowdsourced database within the many police departments in large areas. I was able to work on the mobile version of this application, specifically on the features that allowed police officers to submit requests to add new information on suspects. This feature didn't allow police officers to directly add information to the database, rather they had to fill out a form with certain requirements, that would be authorized by their superiors, to be added to the database. This technical experience goes hand in hand with the STS case being discussed too, describing a software that can be updated to help current society. To work on this application I had to learn how to use the Flutter framework, typescript, and different database languages required to access existing police records. The flutter framework is a web and mobile application software developing kit that uses typescript to develop applications. SOme soft skills I practiced at this internship were best practice communication methods using Slack and MS Teams and presenting at daily standups. My experiences from Program and Data Representation (CS 2150))and Advanced Software Development Techniques (CS 3240) helped a lot with learning how to use these tools. CS 2150 taught data structures which were used throughout this internship and CS 3240 taught web app development and the necessary teamwork skills needed to work on large scale projects like this. CS 3240 also taught us agile methodology, which is a project

management strategy used when designing and implementing software solutions including the one I worked on in my internship. While there were many things that I had already learned in school that came in handy during this internship, there were some shortcomings of my education that showed as well. ONe major omission from our computer science curriculum is application based learning. Many UVA students aren't exposed to popular industry tools until after college, which makes it harder for them to find jobs down the line because experience using these tools are usually requirements for jobs. A way to improve the computer science curriculum is to start introducing these popular tools used in the software development industry earlier on in UVA CS students' academic careers, giving them far more experience and skills earlier in their careers which makes it easier for them to grow in internships and jobs while they are still in college. Another way to make students better prepared for internships like this would be to make co-ops mandatory for CS students for at least 1 of their 4 years at UVA and populating grounds with co-op opportunities to make it easier to find them. This would let students apply what they're learning at school using industry popular tools and introduce students to larger workplace teams and collaboration methods. (Apostolides, V., & Looye, J. W. , 1997). To further research how to improve how our CS curriculum is structured we can look at how other universities that require students to do co-ops structure their curriculum and how they space out courses through a student's 4 years to not cram too much at once. To research this, a set of colleges that do require students to do a co-op or a set of colleges that have an extremely large amount of co-ops available to students would be very helpful to determine how hard it would be to implement this requirement in the UVA CS curriculum.

STS Project

In 2011, the Los Angeles Police Department (LAPD) launched a program called Operation LASER (Los Angeles Strategic Extraction and Restoration). It used previous data on gun-related crimes to map out problem areas and point of interest (Bhuiyan, J. , 2021). These targeted area gets more policing, leading to more data, creating a cycle of policing in these areas. While the technology is meant to help police catch crime, it also does huge social and political work by marginalizing black and latino people in these targeted areas.

Ideally, algorithms are supplied with reliable data from experts so that they can produce predictions that have the highest probability of success. However, recent algorithms have begun to use datasets from the population that have unchecked information, underreported information and more importantly, biased information (Sun, W., Nasraoui, O., & Shafto, P., 2020). Algorithmic risk management is a way for police to determine whether certain locations/areas are low or high risk for crime (Hamilton, M., 2021). This, in theory, is supposed to help police target high risk crime areas before crime occurs so that the police can stop the crime from happening in the first place. However, because of the implications that the predictions make, certain areas and people are marginalized/targeted more by the police. This, in turn, causes an inequality between those that the prediction models assess as high risk vs. those that it assesses as low risk. If we continue to use the datasets that crime prediction algorithms like Operation LASER are based around, we will undoubtedly see a pattern in locations that are deemed "high-risk crime zones".

The issue I am addressing in this paper is the bias produced by Operation LASER by using data from the past. Although the police forces that do use these predictive models don't mean to have biased predictions, they cannot avoid them (Goel, S., Shroff, R., Skeem, J., & Slobogin, C., 2021). This means that the bias is an unintentional effect of the datasets , making it

a prime example of implicit bias. Predictive models use the past to make educated guesses about the future, but when our nation's past has been filled with inequality then it only makes sense that those inequalities will bleed into the predictions of the future. As you could see, because crime Operation LASER continued to use these datasets, currently marginalized races will continue to be marginalized and targeted by the police in the future. This implicit bias in Operation LASER can also be seen in healthcare algorithms (Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S., 2019). Black patients seem to get less healthcare treatment than white people because of the datasets used to tell healthcare professionals where to look when deciding where to put more healthcare efforts. Again, the decision is not intentional, rather the algorithm takes in biased data that makes the decision for the healthcare professionals (Angwin, J., Larson, J., Mattu, S., & Kirchner, L., 2016). Another source of evidence to show that implicit bias in data has a major role in marginalizing certain groups is a study in England and Wales that showed racial injustices were a result of bias in datasets that were used to train machine learning algorithms that predicted crime (Babuta, A., & Oswald, M., 2019). After looking at the corroborating pieces of evidence I argue that the implicit bias that crime prediction algorithms like Operation LASER are based on, results in disadvantages in black and latino communities and advantages in other communities. A way to improve these prediction models would be to use newer algorithms that can detect bias and remove it from the equation when predicting crime (Kim, S., Joshi, P., Kalsi, P. S., & Taheri, P., 2018). Algorithms like k-nearest neighbor, random forest, support vector machine, and LSTM can all do this job and should be implemented in the future (Zhang, X., Liu, L., Xiao, L., & Ji, J., 2020).

Conclusion

My work at TSCTI made me learn a lot of industry popular tools and also let me apply what I learned in school in terms of web application development, agile methodology, data structures and working in a team. I think there are some gaps between what we learn in school and what is used in application in the real world, and solutions like using industry popular tools earlier on in our education, or making co-ops mandatory for students to learn what tools are used outside of college would be a good way to combat the gaps. I was able to help design an app that let police share information more easily. The experience at TSCTI let me learn more about how criminal data is used to police and what data police officers use on a day to day basis. The STS research discussed in this paper aims to show the political effects of implicit bias in datasets used in crime prediction algorithms like Operation LASER. This effect was shown using the theory of technological politics.

Word Count: 1789

References

Mayson, S. G. (2018). Bias in, bias out. *Bias Out (September 28, 2018)*, *128*.

Sun, W., Nasraoui, O., & Shafto, P. (2020). Evolution and impact of bias in human and machine learning algorithm interaction. *Plos one*, *15*(8), e0235502.

Hamilton, M. (2021). Evaluating Algorithmic Risk Assessment. New criminal law review, 24(2), 156-211.

Goel, S., Shroff, R., Skeem, J., & Slobogin, C. (2021). The accuracy, equity, and jurisprudence of criminal risk assessment. In Research handbook on big data law (pp. 9-28). Edward Elgar Publishing

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias risk assessments in criminal sentencing. ProPublica, May, 23.

Babuta, A., & Oswald, M. (2019). Data analytics and algorithmic bias in policing. The royal united services institute for defence and security studies.

Kim, S., Joshi, P., Kalsi, P. S., & Taheri, P. (2018, November). Crime analysis through machine learning. In 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 415-420). IEEE.

Zhang, X., Liu, L., Xiao, L., & Ji, J. (2020). Comparison of machine learning algorithms for predicting crime hotspots. IEEE Access, 8, 181302-181310.

Apostolides, V., & Looye, J. W. (1997). Student assessment of the co-op experience and optimum integration of classroom learning with professional practice. Journal of Cooperative Education, 32, 16-30.

Bhuiyan, J. (2021, November 8). LAPD ended predictive policing program amid public outcry. A new effort shares many of their flaws. The Guardian