Undergraduate Thesis Prospectus

Model Evaluation Service: Improving Machine Learning Model Development Efficiency

(Technical Research Project in Computer Science)

The Struggle over Digital Privacy in the United States

(Sociotechnical Research Project)

By

Alex Kwong

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.

Alex Kwong

Technical advisor: Rosanne Vrugtman, Department of Computer Science

STS advisor: Peter Norton, Department of Engineering and Society

General Research Problem

How can machine learning improve productivity in organizations?

Recent developments in machine learning (ML) offer new possibilities in automation that may transform numerous economic sectors and boost organizational efficiency. Yet, ML also threatens to accelerate problematic trends, including invasion of personal privacy, unimpeded user data collection and monetization, and propagation of targeted misinformation (Kerry 2020).

Model Evaluation Service: Improving Machine Learning Model Development Efficiency

How can better methods of evaluating models improve the efficiency of developing machine learning models?

This capstone project is being individually done under the computer science department with Rosanne Vrugtman as my advisor.

Machine learning is leveraged in almost all Amazon products and services such as Amazon Web Services, Prime Video, and Amazon.com. At Amazon, applied scientists and engineers are tasked with building machine learning models and data pipelines to enable these services. Performance evaluation is a critical step in improving these processes; however, model evaluation is currently being performed manually and individually.

The goal of the project was to build the Model Evaluation Service, a centralized and standardized service for model evaluation. This reduces the number of times models are evaluated and increases development efficiency overall. Rather than improving the state of the art of model evaluation, the goal was to automate that process. In order to do this, automated data pipelines were designed and built to programmatically process incoming model results and produce analyses and visualizations. This was done by using APIs to interact with AWS services such as Lambda and Glue. The end of the internship resulted in a basic implementation of the Model Evaluation Service. The work done only serves as a groundwork for the service and many changes will need to be made because of limitations discovered during the internship.

The Struggle over Digital Privacy in the United States

How do proponents and critics of big data analytics advance their respective agendas?

Big data can improve decision making, model human behavior, reveal new revenue streams, and improve customer experience. However, the collection and use of user data can compromise privacy and encode biases into products and services (Lerman, 2019). The problem is to protect personal data without unduly constraining valuable applications of data analysis in the marketing of products and services.

The struggle begins with companies that collect and use consumer data. Meta, formerly Facebook, is a company that is commonly criticized for its data practices. The nature of Facebook.com as a product means Meta will have data on all of its users from name, age, and other life details. Meta claimed that they do not sell people's data but rather categorize users and employ an ad-targeting system which sells the ability to target and advertise to specific categories to others (Monroe, 2019). Vice president of ads Rob Goldman stated, "You are entitled to your opinion, but we don't sell people's data. Period. That's not a dodge or semantics, it's a fact." Meta makes the distinction that they do not sell private data but instead categories of people that can be targeted by advertisements. A tactic they have repeatedly used is to claim they are working on improving privacy security; however, any improvements are overshadowed by constant privacy breaches and violations (Wagner, 2018). Companies like Meta offer a product like Facebook for “free” when in reality the users themselves are the products (Mian 2016).

The Federal Trade Commission (FTC) is responsible for the enforcement of civil antitrust law and the promotion of consumer protection. In 2019, Facebook was fined $5 billion for deceiving users about their ability to control the privacy of their personal information, violating a 2012 FTC order (FTC, 2019). The FTC imposed new restrictions on their business operations with the goal of preventing future privacy violations. This included an independent privacy committee being instated, removing control by Mark Zuckerberg over decisions affecting user privacy. In the past two decades, the FTC has enforced hundreds of actions against companies for privacy and security violations ranging from sharing health related data to failure of protecting sensitive personal data.

Privacy advocacies such as the Electronic Frontier Foundation (EFF), the Digital Advertising Alliance (DAA), and the American Civil Liberties Union (ACLU) operate in the interests of the public like the protection of their data and privacy. These organizations have a long history of defending and advocating for First Amendment rights in the area of digital privacy. Some examples include the EFF and ACLU challenging the Communications Decency Act of 1996 which resulted in the victory in *Reno v. American Civil Liberties Union* (1997) and challenging the Child Online Protection Act of 1998 which resulted in the victory in *Ashcroft v. American Civil Liberties Union* (2004) (Hudson Jr., 2017). As technologies develop, laws cannot keep up with all the new methods for invasions of privacy. EFF fights in the courts to maintain the digital rights of consumers.

DAA enforces responsible privacy practices while giving consumers information and control over the types of digital advertising they receive. The DAA maintains the Self-Regulatory Principles, a collection of principles that aim to enhance transparency and control for consumers. Compliance with these principles is enforced by the Digital Advertising

Accountability Program (DAAP) and the accountability division of the Association of National Advertisers.

Big data necessitates data privacy protections. Services in which the consumer is the product induce resistance to mass data collection among privacy advocates. Corporations' business interests must be balanced against consumers' data privacy rights.

References

FTC. (2022, January 27). FTC imposes $5 billion penalty and sweeping new privacy restrictions on Facebook. *Federal Trade Commission.*

Hudson, D.L., Jr. (2017, Dec.). Electronic Frontier Foundation. *The First Amendment Encyclopedia*.

Kerry, C. (2020, February 10). Protecting privacy in an AI-driven world. *Brookings*.

Lerman, J. (2019, April 30). Big Data and Its Exclusions. *Stanford Law Review 66*, 55-63.

Mian, A., & Rosenthal, H. (2016). Introduction: Big Data in Political Economy. *RSF: The Russell Sage Foundation Journal of the Social Sciences 2*(7), 1-10.

Monroe, A. (2019, February 15). Fact check: Does facebook sell your personal data? *The Arizona Republic*.

Tarnoff, B. (2018, March 14). Big data for the people. *The Guardian*.

Wagner, K. (2018, September 28). Why should anybody trust Facebook with their personal data? *Vox.*