Thesis Project Portfolio

Optimizing an ETL Pipeline from Multiple Angles

(Technical Report)

The Role of Ethical Design in Building Trust in Machine Learning Algorithms

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Akshay Choksi
Spring, 2024
Department of Computer Science

Table of Contents

Executive Summary

My Capstone project's technical enhancements directly contribute to improving data accuracy, reliability, and real-time access for stakeholders within the bank. Simultaneously, my STS project's focus on ethical design principles ensures that these technological advancements are implemented in a manner that respects user privacy, promotes fairness in algorithmic decision-making, and fosters transparency in data handling practices. As a result, this portfolio provides methodology and research for both projects that contribute value to the machine learning (ML) industry. By integrating technical improvements with ethical considerations, the collective impact of these projects not only optimizes operations but also reinforces the bank's commitment to responsible and trustworthy data-driven strategies. Therefore, this STS Thesis builds a stronger foundation for leveraging advanced technologies, like machine learning, effectively, and promotes harmony between producers and end-users.

The national bank where I served a summer internship was experiencing inefficiency and lack of scalability in its data processing pipeline, hindering timely decision-making in the rapidly evolving landscape of digital marketing. To address this challenge, we leveraged Apache Spark and Delta Lake for data migration, AWS Glue for ETL processes, and designed a robust data fabric architecture for seamless data integration. This multifaceted approach required a combination of programming skills, data engineering expertise, and cloud infrastructure management. As a result of these efforts, we achieved a streamlined Last Touch Attribution process enhancing affiliate payouts. This infrastructure led to improving marketing strategy formulation, model optimizations, and deeper insights into customer conversion journeys. The system now offers greater data accuracy, reliability, and real-time access for stakeholders, demonstrating the importance of data analytical derivatives such as machine learning.

The integration of ethical design principles into machine learning algorithms is a critical endeavor in modern technology development. This research employs the Social Construction of Technology (SCOT) framework to explore the intricate relationship between social factors and technological design in the context of machine learning systems. The central research question of the study investigates embedding ethical considerations into the design process: How can the ethical considerations surrounding machine learning algorithms be integrated into its technological design to increase stakeholder trust? Through a systematic literature review targeting machine learning developers, end-users, and policymakers, the research uncovered insights into the ethical challenges faced during machine learning development helping to understand stakeholder perspectives on these issues. Findings suggested that ethical considerations, including transparency, fairness, privacy, and accountability, are pivotal in shaping the trustworthiness of machine learning algorithms. Moreover, the research identified potential solutions and best practices for addressing these ethical concerns within the machine learning development lifecycle.

Working on both the Capstone project and the STS research project provided me with a unique and enriching experience that extended beyond technical skills. While the Capstone project immersed me in the technical aspects of data processing, analytics tools, and infrastructure optimization, it also sparked my curiosity about how these advancements impact different stakeholders, from decision-makers to end-users. This curiosity was further explored through the STS research project, which allowed me to delve deeper into the non-technical aspects of technology implementation, such as ethical considerations, societal impacts, and stakeholder perspectives.

Through these research experiences, I gained a more comprehensive understanding of the interconnectedness between technology and society. I learned that technological advancements are not just about efficiency and innovation but also about responsibility, ethics, and trust. The STS project provided me with insights into the importance of transparency, fairness, and accountability in technological design, which complemented the technical expertise I developed during the Capstone project. Moreover, working on both projects enhanced my critical thinking skills, as I had to navigate and balance technical challenges with ethical considerations and societal implications. This dual perspective allowed me to see the bigger picture and consider the broader impacts of technology beyond its immediate technical functionalities. Ultimately, the value of working on both projects simultaneously lies in the holistic perspective it provided, integrating technical excellence with ethical awareness to create impactful and responsible solutions in the ever-evolving landscape of digital marketing.