Rules Builder Application for Marketing Campaigns

Improving Consumer Data Collection Privacy

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
Benjamin Ainley

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

Kent Wayland, Department of Engineering and Society

Briana Morrison, Department of Computer Science

General Research Problem

How can user confidence in companies handling their data be increased?

In today's world data is constantly being collected and stored by companies. This can be for several reasons such as personalizing user experience, improving marketing strategy, or selling customer data. Customers may not even know what kind of data is being collected, or feel they have no say in what kind of data large tech companies are collecting. Advances in data collection technology chains have enabled companies to collect, track, and analyze information, patterns, and habits about a user. This form of data is largely unregulated and can be sold to other companies without user consent. Because of this lack of control, customers may feel that their privacy is being violated and lose trust in companies. This data could also be potentially used for unethical purposes causing further harm to consumers or organizations.

Data collection is important for companies, making it easier to understand what customers want and how they interact with the company. By knowing about your customers, a company can tweak the business to better fit needs (Freedman 2022). However, companies that collect user data are burdened with managing and protecting user data at a cost of trust and accountability (Crowcroft 2019). Users expect their data to be well managed and protected. Identity Theft Resource Center published a report indicating that a record number of 1862 data breaches occurred in 2021 in the United States. The previous record was 1506 set in 2017. Most of these breaches occur within the healthcare, finance, business, and retail sectors (Chin 2022). With a growing number of data breaches consumers may begin to lose trust in companies collecting their data, especially because federal law does not protect users from all data collection or the sharing of personal data.

Technical Topic: Creating a Rule Builder App Using React-Awesome-Query-Builder for Marketing Campaigns

How can React-Awesome-Query-Builder be used to create a streamlined interface for building marketing campaigns rulesets?

This summer, I worked as an intern for Capital One in their marketing line of business, working closely with marketers. An important task carried out by these business professionals is the creation of rulesets to filter groups of customers to be targeted for marketing campaigns. These rulesets outline desired specifications of customers to be advertised to. The process of building these rulesets and filtering customers can be tedious work since there is not an app with a streamlined user interface in my line of business. Currently, most of the process is done through command line tools, or two separate applications. These tools are not intuitive and can be difficult for business professionals to use as it requires technical knowledge of query languages. To address this problem, as an intern, my team worked on the development of a rule building application that is easier for business professionals to understand.

To complete the design of the project, my team broke it up into three parts: backend, API, and frontend. First, we worked on the backend, which is responsible for storing the rulesets created by marketers. These rulesets are stored in JSON format, which is a lightweight format for storing data. The tool we used to store these JSON objects is called Amazon S3, an object storage service. Amazon S3 is widely used by enterprises because of its scalability, data availability, security, and performance.

In order for the user to retrieve the data stored in S3 from the website, an API request needs to be made. The tool we decided to use as a communicator between the frontend and backend is called AWS Lambda. AWS Lambda is a serverless, event-driven compute service that runs code in response to events. We decided to use AWS Lambda due to cost and efficiency. It is a cheap option since you only pay for what you use. Price is based on the number of requests and the time of execution. The developer also does not have to worry about provisioning resources, improving productivity, and scaling is handled automatically (Robinson 2020). Because we will be handling millions of customers' records, our computation needs to be scalable and cheap, making AWS Lambda a good choice for the project.

Lastly, we worked on the frontend. The goal of the frontend is to make ruleset creation easy to use and understand for marketers. Business professionals may also not have much experience with query languages, so it is important to make sure that the ruleset interface is displayed in natural language for easier understanding. To accomplish this, we decided to build our frontend using the JavaScript library React, along with a component called React-Awesome-Query-Builder. This React component allows for a user-friendly way to build rulesets (filters). Marketers can use this to customize rulesets by creating fields or groups of fields, and then saving these filters in JSON format. These JSON objects can then be stored in our backend and retrieved later.

This rules builder application is still in the early phases of development and is not yet ready for production. There are different directions the application could head, and has not been fully determined yet. A feature that could improve ease of use would be a better search functionality through the use of tags. On a larger scale, the application needs user privileges, allowing for different levels of authorization in order to increase security. After the end of my

internship, we onboarded my manager's team and they picked up where we left off. They are still working on the app, continuing to improve it. As of right now, they have built their first official rule for the company. The plan now is to enable a user interface for the marketers in quarter two of 2023.

STS Topic: Policy Response Analysis of the Facebook – Cambridge Analytica Data Scandal

How have different state and government policies been affected by the Facebook – Cambridge Analytica Scandal?

In 2015 Cambridge Analytica was able to gain access to users' personal data by leveraging its alliance with Facebook. This user specific data was combined with psychological profiles developed through a personality quiz taken by users, giving Cambridge Analytica the ability to target users with messages that could possibly influence behavior (Isaak 2018). This scandal played a large role in the erosion of trust in consumers in companies. Profiles of users can be made, and users can be subconsciously influenced through messages without their consent. These users implicitly expect that users of their data will respect their privacy. With a growing number of ethical issues linked with data collection and access, trust in companies is being lost (Miltgen 2019). There are so many instances of ethical issues related to data collection due to the personal data ecosystem being mostly de-regulated, fragmented, and inefficient. Users are usually not given control over their personal data, leading to issues related to privacy, personal data ownership, and transparency (Crowcroft 2019). These issues are constantly being discussed, leading to different policy ideas and responses. Looking into different policy

responses of the Cambridge Analytica scandal can give a better understanding on how to improve consumer privacy and trust, while still allowing companies to use personal data.

This scandal began when Global Science Research (GSR) initiated a research project in cooperation with Cambridge Analytica in order to identify the parameters necessary to develop "OCEAN" psychological profiles. These profiles were developed using a personality quiz and required users to allow access to their Facebook profiles, which gave the company access to the participants' friends' data through the Facebook Open API. The goal of the research was to establish a methodology for psychographic profiling of individuals; therefore, it was not necessary for the company to keep specific user data to conduct their research. However, Cambridge Analytica realized they could combine these user profiles with other data from social media, browsers, online purchases, voting results, and more. With all these data points, users could be micro-targeted with messages or advertisements likely to influence their behavior (Isaak 2018).

Following the Facebook – Cambridge Analytica case, the FTC imposed a $5 billion penalty and new privacy restrictions. This was the largest penalty ever imposed on a company for privacy violations, with the second greatest being the Equifax data breach being $275 million (Federal Trade Commission 2019). This highlights the significance of this scandal with the fine being 18 times greater than the Equifax data breach. With millions of US citizens using Facebook daily, Facebook has a large responsibility to protect their user's data. On top of this fine, the FTC imposed a new privacy compliance system for Facebook. This order creates greater accountability by establishing an independent privacy committee of Facebook's board of directors. Facebook will also be required to designate compliance officers responsible for their privacy program. Additionally, the order calls for a significant number of new privacy

requirements which ensure security and prevent collection of certain user data (Federal Trade Commission 2019).

Surveys after the Cambridge Analytica scandal indicate that customer trust in Facebook has dropped by 66 percent. A year before the scandal, 79 percent of users believed that Facebook was committed to protecting the privacy of their personal information. This dropped to 27 percent immediately after the scandal, and 28 percent after Zuckerberg's testimony with Congress (Weisbaum 2018). Users wish to have more control over their data and want companies to disclose how it uses the personal information that is collected.

This scandal had a massive impact worldwide with governments of various countries all investigating the data breach. It also may have significantly impacted the California Consumer Privacy Act (CCPA) and the EU's General Data Protection Regulation (GDPR), two significant pieces of privacy legislature which provide comprehensive consumer privacy laws.

The California Consumer Privacy Act is a state statute intended to enhance privacy rights and consumer protection for residents of California, and is the United States' first comprehensive data privacy law. The statute was originally hastily passed and put on the November 2018 ballot. It is the toughest privacy law in the United States, which slightly concerned lawmakers. They initially felt that this law was too consumer privacy focused instead of striking a balance between consumer protection and preserving innovation. However, the Cambridge Analytica scandal caused the ballot initiative to gain more traction, and the California legislature faced pressure from data privacy advocates in the states of Washington and California. The CCPA then went into effect in January of 2020 after numerous amendments (Lee 2020).

The GDPR was approved by the European Union in 2016 and went into effect in May 2018. Although this is an EU regulation, the territorial scope of it applies broadly. The GDPR

applies to any business that processes personal data of subjects in the European Union. Therefore, if a company based in the United States collects data from anyone in the EU, the GDPR applies (Lee 2020). The GDPR is an influential piece of legislature, and its awareness was heightened due to the Cambridge Analytica scandal. With regulators worried about the adequacy of security and data protection controls, it is likely that it had an impact on the way GDPR controls will be enforced and implemented. The scandal may have accelerated investigations and enforcement actions from the European Data Protection Authorities that may have otherwise been implemented with some restraint (Simberkoff 2018).

Concerns of privacy grew in other parts of the world as well after the incident with Cambridge Analytica. In India, firms and the people have become more aware of their data and factors related to privacy. Because of this, the Indian Courts agreed that a right to privacy is a fundamental and should be considered as one of the fundamental rights included in their Constitution. The uproar among the Indian population has led to the creation of the Data Protection Bill in 2019, defining the rights of data principles, the obligations of data fiduciaries, and penalties for non-compliance. This bill took inspiration from the EU's GDPR with the goal of giving users more control over their personal data (Verma, Jawanda, and Kaur 2021). Although it is still in the works, this bill highlights the growing desire for data privacy and control.

For my research I will be conducting a case study on the Facebook – Cambridge Analytica scandal, analyzing all the actants involved in the situation, such as the people affected and how different policy makers responded. A large data breach like this affects many people, causing a reaction by the public and governments. Information will be gathered through legislative hearings and law review articles regarding the Cambridge Analytica scandal, along

with responses from the public in different parts of the world. These responses to the scandal have caused for reform in policies related to personal data collection and use by companies. I plan to survey the different policy responses to this scandal and compare them in order to better understand the range of different values that informed these responses.

Conclusion

Through my STS research project, I hope to gain a better understanding of how the collection and sharing of personal data is regulated and how it affects consumers. Currently, the US data ecosystem is largely deregulated and consists of a patchwork of federal and state laws. By looking into how the system is set up, along with proposed policies, a better understanding of how personal data collection and sharing can be reached. I have had some exposure to how customer data is handled while interning at Capital One while working in the marketing line of business and hope to dig deeper into how this data is used by companies. Consumers may feel like they have a lack of control of how their personal data is being used or may feel that their privacy is being violated, causing a loss of trust in companies. By performing this proposed research, hopefully a balance between data collection and invasion of privacy can be reached, increasing user confidence in companies handling their data.

References

Chin, K. (2022, August 5). *Biggest data breaches in US history* [blog post]. Retrieved from https://www.upguard.com/blog/biggest-data-breaches-us

Federal Trade Commission. (2019, July 24). *FTC Imposes $5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook* [Press Release]. Retrieved from https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions-facebook

Freedman, M. (2022). Businesses are collecting data. how are they using it? *Business News Dail*y. Retrieved from https://www.businessnewsdaily.com/10625-businesses-collecting-data.html

Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*, *51*(8), 56–59. https://doi.org/10.1109/mc.2018.3191268

Lee, C. (2020). The Aftermath of Cambridge Analytica: A Primer on Online Consumer Data Privacy Note. *AIPLA Quarterly Journal*, *48*(3), 529–568. https://heinonline.org/HOL/P?h=hein.journals/aiplaqj48&i=544

Miltgen, C. L., Cases, A.-S., & Russell, C. A. (2019). Consumers' responses to facebook advertising across pcs and mobile phones. *Journal of Advertising Research*, *59*(4), 414–432. https://doi.org/10.2501/jar-2019-029

Pal, R., & Crowcroft, J. (2019). Privacy trading in the surveillance capitalism age *viewpoints on 'privacy-preserving' societal value creation*. *ACM SIGCOMM Computer Communication Review*, *49*(3), 26–31. https://doi.org/10.1145/3371927.3371931

Robinson, D. (2020, December 9). *Serverless: Weighing up the pros and cons for enterprises*. ComputerWeekly.com. Retrieved from https://www.computerweekly.com/feature/Serverless-Weighing-up-the-pros-and-cons-for-enterprises

Simberkoff, D. (2018, August 30). How facebook's Cambridge Analytica scandal impacted the intersection of privacy and regulation. *CMSWire.com*. Retrieved from https://www.cmswire.com/information-management/how-facebooks-cambridge-analytica-scandal-impacted-the-intersection-of-privacy-and-regulation/

Verma, A., Jawanda, K., & Kaur, A. (2021). Data Privacy and Cambridge Analytica: A Case Study. *Supremo Amicus*, *24*, 368–380. https://heinonline.org/HOL/P?h=hein.journals/supami24&i=368

Weisbaum, H. (2018, April 18). *Trust in facebook has dropped by 66 percent since the Cambridge analytica scandal*. NBCNews.com. Retrieved from https://www.nbcnews.com/business/consumer/trust-facebook-has-dropped-51-percent-cambridge-analytica-scandal-n867011