Thesis Project Portfolio

Working at a Start-Up: Experiences and Learnings as a Data Science Intern

(Technical Report)

A Gloomy Cloud: A Socioeconomic Analysis on the Use of Data Centers in Northern Virginia

(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

Ansel Sanchez

Spring, 2023

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

With the advent of more advanced internet infrastructure, data has become ever more important. Because of this increased data reliance, I aimed both my technical and STS research projects on thinking about the long-term sustainability of data. For my technical project, I developed a robust full-stack data monitoring, visualizing, and alerting system for engineers to use at a local hydroponic start up. In my STS research project, I investigated the effects of data centers on the Northern Virginia community and how these facilities can be integrated better. In both of these projects, I was able to develop creative solutions in creating an environment for better data usage, whether that be in a more technical one to improve data accessibility and visibility for engineers or in a more societal one to improve data center presence in regions that feel threatened by its increasing prevalence.

For my technical project, I worked at a local Charlottesville start-up, Babylon Micro-Farms, which specializes in developing smart cabinets for automated hydroponic produce farming. The company needed a new system for better monitoring of its incoming data from their hydroponic farms. As a data science intern, I was tasked with building a new dashboard of sensor metrics to improve how engineers would see their data. On top of this, it had to be easy to use and aesthetically pleasing enough to be presentable to potential investors. Working in an Agile environment to complete the project, I used languages and technologies I was familiar with such as Python, Linux, and version control on top of a myriad of new tools learned throughout the internship experience. In building the system, thinking about how the application could be sustained past my internship was a key priority, mainly thinking about how it would be sustainable for continued development with future engineers. This included building it well enough to withstand changes in data format, data ingestion, and other backend alterations the

engineers could perform. In the end, I developed, for both internal and client use, a full-stack data scraping, visualizing, and alerting application to fully monitor all of the company's farms while prioritizing longevity and robustness to ensure successful future use of the system.

As a way to think about how data and its related technologies can be better for society and endured usage, I focused my STS project on analyzing the effects of data centers on the lives of residents in Northern Virginia, one of the most prominent places of data center construction. Also motivated by my own experience living in the area and hearing complaints about data centers encroaching on the community, I aimed my project on interviewing community members, both those involved with data centers/related technologies and those that aren't, to understand how they perceived data centers being built in their backyards. At the conclusion of the project, I found that these sprawling facilities were seen to be positive in that they bring jobs to the area and stimulate the economy, however, are greatly detrimental to the community by overusing natural resources, driving up living costs, and ruining the local suburban culture and aesthetics. With understanding what effects data centers have on the community, I also understood how they can be better integrated in their human environment, which included converting abandoned buildings into data centers and using more discreet facades to better blend in with the area.

By accomplishing projects with sustainable data and technological uses in mind, with one in the technical sphere and one more in the community-oriented sphere, I really got to understand how the two seemingly disconnected worlds could actually go hand-in-hand. For example, just seeing the sheer amount of data there is, even for a small start-up company like Babylon, worsened my concern that data and internet operations are a taxing endeavor for data storage and processing facilities, which partly inspired my investigation on how to improve data

center integration. Furthermore, one key principle of software design is creating systems that can be maintained for future and continued use. This was an integral part of my technical project and one that made me think about how it could be applied to non-technical areas, such as with local community integration of data centers. With internet needs becoming ever more apparent, data centers will become a staple in many other areas of the country, so thinking about how the construction, operation, and continued use of them can be better embedded in residential areas for future use is an important concern to ensure the relationship between technology and community continue to healthily coexist together.