Gabbie Bolcer

Portfolio Essay

In my portfolio I included works to demonstrate ways I have done something that is very important to me, which is creating meaning from data. I included two works from my Computer Science classes that are common implementations to create meaning from data and a visualization I created in one of my statistics courses. The What's for Dinner Assignment uses machine learning to figure out patterns in the data and derive meaning from the dataset set. The Two Way Directed Graph implements graph theory which is used in modern APIs to figure out how different people, events, or attributes are connected. Additionally, my Copy the master's assignment derived meaning from data by creating a visualization that people outside of the data world could understand. The problem of interest I will be exploring in this essay is deriving meaningful insights from the vast amount of data in the modern world.

We are at a phase in the digitized world where we are creating more data and datapoints than we know what to do with or have time to compute. Many of our basic day-to-day modern tasks leave mountains of data exhaust. Data exhaust is data generated as a byproduct of people's daily actions, often without them realizing it. For example, if you used the internet today, every site you visited and how much time you spent on that site is an example of data exhaust. You didn't intentionally give this information to anyone, it just happened to be a byproduct of the products you were using and the capabilities of those products. Many of these points of data exhaust are collected and put into large data sets that companies later use later to better implement their product. This is how a lot of the "Big Data" you've heard about comes into existence.

Many modern digital products use machine learning to improve their product. An example of machine learning in digital products could be using machine learning to create the best streaming experience for users, buying and selling stocks based on previous and current market trends, or increasing engagement on social media platforms. While the machine learning algorithm I implemented in my What's for Dinner assignment doesn't answer any big business or real-world questions, it was a good introduction to how machine learning works behind the scenes. During development, I had to figure out how the algorithm was going to calculate which attribute (column in the dataset) was going to be the most important in deciding the output variable. I also had to figure out how to filter out the attributes that didn't affect the outcome which was where someone decided to eat for dinner. This assignment helped me understand how machine learning works and some of the basics of how its applied to solve big questions in the real world. Additionally, I got firsthand experience in understanding why machine learning algorithms make the decisions they do. In understanding why machine learning algorithms make the decisions they do, I hope to be able to detect bias in algorithms that further our current state of the world instead of improving it.

This assignment made it evident to me how we can use machine learning to understand trends in big datasets in the real world. During my time at St. Thomas, I took many Statistics courses. Most upper-level statistics courses consisted of a final project at the end of the semester to demonstrate the statistical techniques learned in the class. These projects generally consisted of finding a dataset and selecting variables to include in a model. Depending on the dataset, it could be very difficult to choose which variables were significant and significantly affected the dependent variable. Implementing machine learning as I did in this assignment looks to solve that problem. The machine learning algorithm can pick up significant variables in a way that

humans cannot (believe me, testing that algorithm to see if it was working was extremely difficult!). Implementing machine learning to understand datasets can help us derive more meaningful and relevant insights from data that humans may not have been able to see on their own.

My Two Way Directed Graph assignment uses graph theory to solve a problem of trying to figure out if you can get from one place to another following the graph paths. The assignment laid out a few additional parameters that the path must follow. This project implements graph theory which can be used to find connections between datapoints and derive insights from data that has direction associated with it. Modern implementations of this can be found in many social media networks' algorithms to figure out who might know someone else. Additionally, graph data structures may be used to generate the recommendations pages that are common on e-commerce websites. Without these data structures and theory, much of the data that companies have used to generate more business and revenue would be useless.

This assignment helped me gain understanding about how data is connected and how meaningful information and insights are drawn from datasets that have connected datapoints. This assignment was a basic implementation of graph theory concepts and did not attach any real-world examples, but it was interesting to the basics behind how algorithms can extract meaning from data that has directionality with it. There is data with directionality all around us. After completing this assignment, I have another type of data that I can work with and provide meaningful information from.

Once someone has derived meaningful insights from a dataset, it is very important to be able to effectively communicate your findings. One of the most powerful tools to showcase findings in your data is a visualization. A visualization needs be as simple as possible to not

confuse the reader, use colors that are easy to distinguish from one another, and implement appropriate graphing methods for the data being described. The art of visualizing data to make it understandable to someone not familiar with the dataset and with a lower data literacy level than yourself was a skill I learned and demonstrated in my Data Communication and Visualizations class. In my "Copy the Masters" assignment I recreate a visualization the was in a piece of published data journalism. This project displays another skill I have acquired during my education that contributes the topic of interest in this paper, which is deriving meaningful insights from data.

I can use my skillsets in generating meaning from data to be a morally responsible leader who thinks critically, acts wisely, and works skillfully to advance the common good as the St. Thomas mission statement says the institution hopes to shape students. In my future jobs and projects as a software engineer who focuses heavily on the data side of the business, I will use the knowledge I gained in my computer science classes to work skillfully and create creative, unique, and efficient solutions to software and data problems I encounter in my job.

As important as it is in my field to be a good programmed and good at manipulating and gathering meaning from data, it is also very important that I can apply critical thinking to the projects I'm working. From this I can access if what I'm doing advances the Common Good. Advanced algorithms and data mining techniques *can* be used to produce a better, safer, healthier, and more equitable world. Unfortunately, employing technical solutions doesn't always result in this better world. Understanding why these algorithms and data analyses fail the way they do is usually a difficult process to understand, but very important so we don't implement solutions that only solve problems for a certain group of people. Using what I learned in my many humanities courses at St. Thomas, I have gained a much deeper understanding of

what the Common Good is for not just myself but also those around me who may not appear to share many similarities with me on the surface. I can use this understanding to act wisely and ensure that the projects I'm working on advance the Common Good.

Once I have identified that my projects align with my learned understanding of the Common Good, I can use my skills to create usable data that I (or another team, if I am working in a group) can create powerful insights from to improve our world. It is very important to me that I contribute meaningful work that improves the world and allows others to be the best at what they do.