

Coleman Lyski

Professor Fred Annexstein

Computer Science – Senior Design

15 September 2020

### Individual Capstone Assessment

Sentiment analysis is used to interpret and classify emotions within textual data using text analysis techniques. My senior design project is a web application that will allow users to search for any set of keywords. The application will then look for the use of these keywords in various online sources, such as social media (Twitter, Facebook, etc.), news outlets (New York Times, LA Times, etc.), and forums (Reddit, Quora, etc.). Finally, the application will perform sentiment analysis across all of the gathered text in order to generate a final report. The goal is to create a report that accurately represents what is currently being said about the searched for keywords across the internet at that given moment.

While my entire college curriculum will guide the development of this project, there are a few courses that stand out from the rest. In order to perform accurate sentiment analysis, tons of textual data is required. Anytime an application is dealing with large amounts of data, it is critically important to make this data as small, quick, and easily accessible as possible. My Data Structures (CS 2028) and Algorithms (CS 4071) courses gave me the knowledge I need to create useful structures for storing and manipulating the textual data. Another helpful course is the Cloud Computing (CS 5165) course I am taking this semester. Powerful computers are going to

be needed to perform the analysis quickly on large amounts of data. Our team will be able to leverage machines on the cloud to do the analysis in a fast and efficient manner.

I have received plenty of experience with building applications from the ground up and deploying them through all my co-ops. On my third co-op with Teradata as a Software Engineer Intern, I was responsible for creating new Angular web applications and getting them deployed to our internal environment. This gave me experience in creating a CI/CD pipeline for a brand-new application, which pretty much mirrors what we will have to do with this senior design project. On my fourth co-op with Siemens as a DevOps Intern, I had to create another web-based application that performed text analysis on large amounts of JavaScript project files. Although the analysis I was doing was not quite sentiment analysis, many of the same techniques I used can be carried over to this senior design project – things like parsing the text, cleaning the text, and storing the text in an efficient manner.

One of the aspects of this project that is the most attractive to me is how many different use cases it can be applied to. Although our team has a solid idea of what we want to create, I believe this project can grow and transform as we work on it. What we are really creating with this project is an extremely solid platform that can then be added to and modified to fit all sorts of different needs or goals. At the core, all this tool will do is grab the real-time sentiment of any keywords across the internet, but what can be done next with the data is the promising part. For example, if one were to start gathering the sentiment of stock tickers and comparing it to the prices of stocks, then this application could be utilized to create an advanced financial tool.

Because this project is so fluid, I think the best approach is to start small and figure out what works and what needs to be rethought. We may want to start with gathering the textual data from a small group of sources – possibly just Twitter and Reddit in the beginning. Then with this text data, we can perform the sentiment analysis to generate a simple report, one that might only give a “positive”, “neutral”, and “negative” output. Finally, we can implement a barebones web application that allows a user to search for what they want and receive the analysis real-time in their browser. Once we have this initial application up and running, we will be able to add in more depth and complexity one piece at a time, such as more detailed reports, drawing from a wider group of sources, and more complex analysis tools.

We will know this project is done when a user is able to access our website, search for whatever keywords they would like, and received real-time sentiment analysis of those keywords from across the internet. We will know the project does a good job when after receiving the sentiment report, the user can perform a Google search with the same keywords and receive articles/blog posts/social media feeds/etc. that share the same sort of general sentiment as the one we were able to infer from our application.