|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | ID | Section Contributed | Section Edited | Other Contributions |
| Thompson, Josh | 301421008 | Written component |  |  |
| Ty, David | 301334955 | Coding component |  |  |

**Assignment 2 – Group 9**

**Description of Program**

The goal of our program was to conduct a sentiment analysis on reddit comments from various posts. We chose to look at comments that were sorted into the controversial category underneath the reddit posts. Controversial comments on reddit are essentially comments that are downvoted and hidden from most users. Our belief was that these comments were downvoted because they contained negative sentiment. Comments were gathered by scraping various posts, while sentiment analysis was conducted through a reimplementation of nltk Vader.

As noted, the input was gathered through scraping reddit posts. These overarching topic of these reddit posts was world news. Specifically, we chose from 5 posts with topics on a Harriet Tubman monument, Saudi Oil, Charges dropped in a road rage incident, Silicon Bank, and Norfolk. From these posts our scraper gathered comments that had been sorted into the controversial category, with all but 1 of these posts resulting in 10 comments each. The scrape on the reddit post regarding Norfolk had only had 1 controversial comment at the time of scraping. Likewise, the scrape from the Harriet Tubman monument post contained 1 deleted comment, for which sentiment could not be conducted on. Resulting in 40 “controversial” comments for our program to conduct sentiment analysis on.

By reimplementing NLKT Vader, we were given the categories of positive, neutral, or negative. Each individual comment was given a sentiment for each of the 3 categories, resulting in a summary of either being positive, neutral, or negative. At the same time each of the 5 topic categories receives a percentage for each category.

[THIS SECTION WILL BE ON OUTPUTS]