

Individual Final Report for Jonathan Schild

Team 3 sought to emotionally classify Reddit comments and posts over time to enable administrators and moderators to track changes in the sentiments of their online communities. Overall, tasks were well distributed and each team member did a good job collaborating and taking responsibility for various parts of the project. While this lack of dividing up subtasks was good for combining knowledge, skills, and expertise, it also created redundancies in the code and functions as individual contributions often overlapped with others' contributions elsewhere. This made cleaning up final files tedious and potentially damaging to the final application.

Although goals and objectives were not initially allocated to individuals, each team member took on responsibilities as time went on and milestones were achieved. I initiated idea generation by pitching several ideas, including an initial preference for something related to mis/disinformation. After combing for datasets and discussing as a team, we settled on tracking sentiments over time for online discussion forums. To get started as early as possible, I developed and trained the team's first Bert model, building it from the ground up with the assistance of online tools, like ChatGT and Claudi.ai. This initial model established the base structure that was implemented in our other models.

We realized that training our models on Twitter data when we intended to implement them on Reddit might pose problems and reduce confidence in their output. I decided to try and use pseudo labeling to add Reddit data into our initial Twitter training data that would be used to train a second BERT-based model. This second model was similar to our initial BERT model with the exception of being trained differently. Finally with two models ready, I wrote the team's code that takes an input model and gets the predictions that that would eventually illustrated in our final application.

Finally, while other team members focused on the scraper, visualizations, and the implementation of Streamlit, I got started on the team's final paper and presentation. I crafted the individuals presentation slides, and drafted our first two versions of the final report for group input, editing, and finalizing. From a coding perspective, we wrote our own scrips, improved and built upon each others, and sought assistance from online tools such as ChatGPT, Claude.ai, and copilot when and where necessary. Measuring specific lines of code across all of our scripts when each team member frequently made edits across all of them is difficult.