**Gauging Changes in Sentiment Using Emotional Classification**

Reddit is a popular social media forum that attracts significant numbers of users, content, and engagement from around the globe. Its ever-growing volume of online community discourse presents a platform that can inadvertently shape opinions, damage brands, and incite real-world action. Identifying and understanding swings in user sentiments is becoming more and more critical to combat inaccurate and inadequate information and general community divisiveness. Our application allows subreddit moderators to monitor their online communities for significant swings in sentiment that could have broadly negative implications if left unchecked. OUTLINE OF REPORT.

We used two datasets to train our final model(s). Our initial models use a dataset from Hugging Face that contains approximately 90,000 tweets that are labeled according to evenly distributed emotions: sadness, joy, love, anger, fear, and surprise.[[1]](#footnote-1) We also use a dataset available on Kaggle that contained almost 8,500 unlabeled Reddit comments from various subreddits.[[2]](#footnote-2)

We experimented with multiple models during our analysis all of which implemented fine-tuning on learning rates, batch sizes, dropout rates, and decay rates. Each model also used the Adam optimizer with the sparse categorical cross entropy loss function and monitored validation accuracy for early stopping.

* In our project, we have implemented a Reddit scraper utilizing the Python Reddit API Wrapper (PRAW) to dynamically fetch and aggregate Reddit posts and comments based on specific user inputs. This tool is adept at retrieving real-time data from designated subreddits, tailored to user-defined criteria. Users can specify the subreddit from which they wish to extract information, ensuring targeted data collection. Additionally, they can define the time interval for the data retrieval, choosing from daily, weekly, or monthly options, which aligns the data collection with their temporal analysis needs. The tool is also equipped to handle requests for a specific number of top posts from the chosen subreddit, with further refinement available through filtering options that sort these posts based on their popularity over the last week, month, or year. Moreover, for each of these top posts, users can define the number of associated comments they wish to retrieve, enabling a comprehensive view of both the posts and the community interactions around them.
* Our first model (model1) consisted of a standard Bert-based multi-class classifier, trained on our initial Twitter dataset using training, validation, and testing splits. BERT allows us to understand the context of a word based on all its surroundings (both left and right of the word). This capability is particularly useful in capturing the contextual nuances of social media language, which is often informal and idiosyncratic. This initial model achieved validation accuracy and F1 score of .95, each. However, since we intended to apply the model to Reddit data, we wanted to introduce Reddit comments into the training dataset, so the model is sensitized to Reddit comment expression and structure. To get Reddit data with labels that matched our Twitter dataset, we applied our initial model to unlabeled Reddit comments that predicted scores for each of our six emotion labels. We then removed comments that did not contain at least one emotion with a score higher than .6 to increase our confidence that the comments contained at least one detectable emotion. These now-labeled Reddit comments were added to our labeled Tweets to create a new combined dataset. After training a similar standard Bert-based multi-class classifier, this new model (model2) achieved a validation accuracy and F1 score of .91, each.
* The second model (ROBERTA) consisted of base Roberta sequence classifier trained on the same twitter and reddit data. We were able to achieve a validation accuracy of 0.9502 and F-1 Score of 0.9502. We chose to train ROBERTA for its optimization on BERT’s pre-training procedures. ROBERTA trains on larger mini-batches and learning rates and removes the Next Sentence Prediction (NSP) objective used in BERT, focusing solely on more extensive masked language modeling. This makes ROBERTA more effective in understanding complex sentence structures, which is common in Reddit comments as users often express detailed opinions and narratives.
* The third model ELECTRA focuses on distinguishing between genuine and artificially replaced tokens in text. This method allows ELECTRA to learn more granular and nuanced patterns of language usage, particularly adept at identifying the subtle linguistic cues often found in social media. Unlike BERT and ROBERTA, which primarily learn from masked words, ELECTRA's discriminative training targets the entire vocabulary, giving it a more comprehensive understanding of language context and structure. This aspect is especially valuable in social media sentiment analysis, where emotions and opinions are frequently expressed through intricate and varied linguistic expressions. By capturing these finer details more effectively, ELECTRA could potentially offer more accurate and context-sensitive sentiment predictions making it a more suitable choice for the complexities and nuances inherent in social media texts. Our model consisted of an ELECTRA small-discriminator sequence classifier achieving a validation accuracy of 0.9531 and F-1 score of 0.9530.
* we have chosen to train these three transformer-based models to get an overall picture of what the patterns or representations these models learn and how these learned representations affect the predictions and sentiment trends of each specific subreddit aiding in our model selection.

Observations

* Mention the max\_length issue that could limit our models. The preference over emotions is captured more effectively in Electra based on the same context on which Roberta and BERT have been trained. However, bert and Roberta could not pick on these differences. All three models captured the trend of each emotion well.

6. Results. Describe the results of your experiments, using figures and tables wherever possible. Include all results (including all figures and tables) in the main body of the report, not in appendices. Provide an explanation of each figure and table that you include. Your discussions in this section will be the most important part of the report.

In summary, we trained two models to predict the emotion and general sentiment of Reddit posts/comments over time, and created an application the displays the change in attitudes over a specified time frame. This tool would offer the most value to administrators and moderators of online communities where text-based communication is both frequent and engaging. They would be able to regularly monitor for abrupt changes to typical behaviors and views in discourse that could threaten the integrity of the subreddit and its reputation. Looking forward, our model would benefit from additional training on more Reddit-specific data. As is, Tweets dominated the training process which could have negatively impacted the model’s predictive power on

8. References. In addition to references used for background information or for the written portion, you should provide the links to the websites or github repos you borrowed code from.

Streamlit?

1. https://huggingface.co/datasets/philschmid/emotion/tree/main/data [↑](#footnote-ref-1)
2. https://www.kaggle.com/datasets/prakharrathi25/reddit-data-huge/data [↑](#footnote-ref-2)