Online Community Monitoring

Tracking Changes in Subreddit Sentiments

Thomas, Jonathan, and Ashish

Group Final Report

Introduction

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 stake holders to monitor their online communities for significant swings in sentiment that could have broadly negative implications if left unchecked.

Our initiative leverages live data extraction from Reddit, gathering posts and comments based on either predefined datasets or user-provided criteria. We employ sophisticated transformer-based models such as BERT, RoBERTa, and Electra, which have been meticulously adjusted for the task of detecting and forecasting sentiment pertaining to six distinct emotions: sadness, joy, love, anger, fear, and surprise. We compile these sentiment evaluations to construct a detailed sentiment trajectory for each emotion over a chosen period. The outcome of this assessment is depicted in a detailed chart that illustrates the fluctuation of these emotional sentiments over time, providing critical insights into the dynamics of subreddit communities.

Description of the data sets

We used two datasets to train our final model(s). Our models primarily use a dataset from Hugging Face that contains approximately 90,000 labeled tweets distributed evenly across six emotions: sadness, joy, love, anger, fear, and surprise.[1] We also use a dataset available on Kaggle that contained almost 8,500 unlabeled Reddit comments from various subreddits.[2] This dataset was used to experiment with pseudo-labeling and incorporating confident test data into our initial Twitter training data.

Description of the NLP models

We have fine-tuned the three different pre-trained transformer networks from hugging face during our analysis out of suspicion that predictions of emotions and sentiment could vary depending on the type of model implemented. With each model we tuned hyperparameters like learning rate, dropout rate and batch size. All of them also used the Adam optimizer with the sparse categorical cross entropy as loss function and monitored validation accuracy for early stopping.

We developed two BERT-based multi-class classifiers for emotion detection in text, as it learns contextual word representation beneficial for informal language. Model 1 (BERT1) was trained on Twitter data split into training, validation, and test sets. BERT1 achieved validation accuracy and F1 of 0.95. To improve Reddit generalization, we applied BERT1 to unlabeled Reddit comments to pseudo-label high confidence examples (threshold 0.6/label) and combined these newly labeled samples with the Twitter training set. Using this semi-supervised dataset, we trained Model 2 (BERT2) with the same BERT architecture. Incorporating Reddit data reduced performance (validation accuracy/F1: 0.91) but improved applicability to this target domain.

Experimental Setup

Our experimental setup involves leveraging a dataset comprising 90,000 tweets, each labeled with one of six distinct emotions, along with subreddit data to train our sentiment analysis model. To evaluate the model's performance, we employ F1 scores for individual emotions, overall accuracy, and a composite F1 score that encapsulates the model's average effectiveness across all emotions. These metrics provide a quantitative assessment of the model's precision and recall capabilities. As there are six emotions to predict, we are going to use the sparse categorical cross entropy loss to minimize the error. Furthermore, we conduct both quantitative and qualitative comparisons to ensure a thorough analysis. Numerically, we measure the model's performance by comparing the calculated metrics against benchmarks of other transformer models. Qualitatively, we examine the trends in the subreddit data to verify that the model's sentiment predictions align with the actual discourse and emotional context present within the data. This dual approach allows us to validate the model's accuracy and its practical applicability in interpreting and reflecting the nuanced sentiment trends within the sub reddit communities.

Implementation and the Execution Flow

We utilized a RoBERTa sequence classifier for our second model, motivated by its optimization of BERT's pre-training procedures and ability to comprehend complex syntactic structures prevalent in Reddit comments. Solely trained on the Twitter corpus, it achieved validation accuracy and F1 scores of 0.95. To diverge from BERT-based models, we trained an ELECTRA model capable of learning more granular and nuanced language patterns ubiquitous in social media. Unlike BERT and RoBERTa which rely on masked language modeling, ELECTRA is context-sensitive and uses token replacement to increase the performance, enabling more precise sentiment predictions suitable for the intricacies in this domain. Our ELECTRA model obtained a validation accuracy and F1 score of 0.95, comparable to our BERT and RoBERTa models.

We used Streamlit to apply our models and display visualizations that help a user identify and understand changes in sentiments of online discourse, in this case subreddits. A scraper using PRAW fetches Reddit posts in real-time based on inputs provided by the user, such as subreddit, number of posts and comments, and date-range criteria. The Reddit API limits scraping to the past day, week, month, or year, instead of providing the ability to specify a date range. to ensure we could measure changes over time, we pulled the dates for each post and comment and later grouped them according to a daily, weekly, or monthly intervals extending out from the present. This allowed us to average sentiments together over the specified interval and visualize the change over time.

After pulling in the live data, our application displays a histogram that allows the user to gauge the number of comments and posts per interval to ensure certain time periods do not contain missing data and that they do not contain and significantly more or less data than other periods. We then provide the
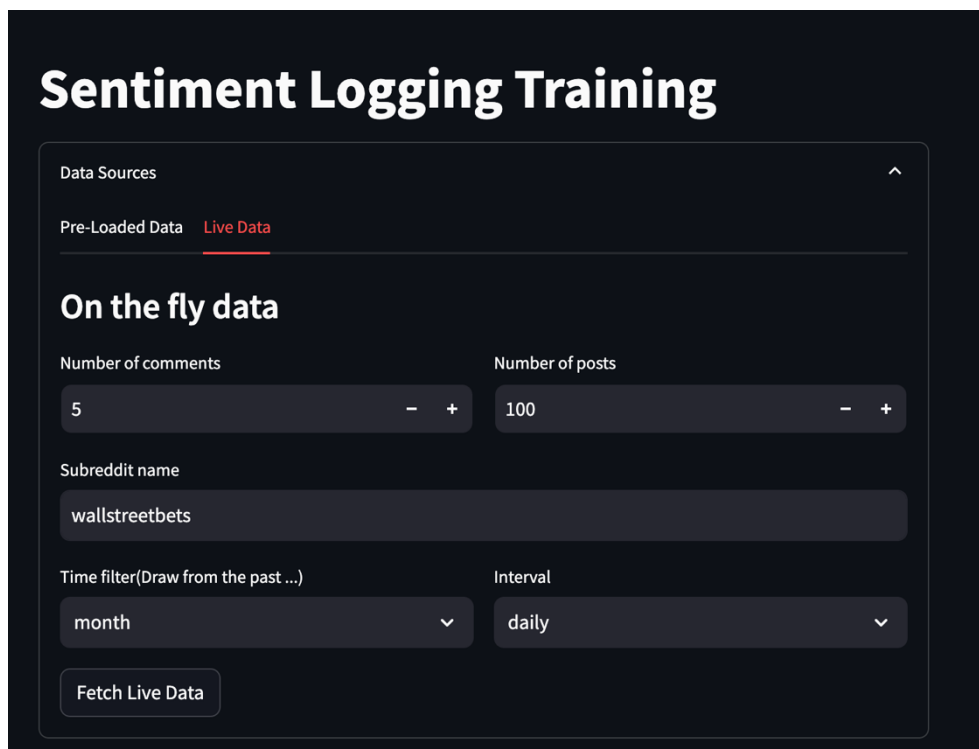
user the ability to select and apply the data using our four models (BERT1, BERT2, RoBERTa, ELECTRA). The application returns two charts: one showing the six emotions plotted on a line chart by time-interval averages, and a second chart showing the positive and negative sentiment over the same time-interval averages for easy evaluation.

Hyperparameter Search

In the fine-tuning of our models, we conducted a hyperparameter search focusing primarily on the learning rate and batch size, two crucial aspects that significantly influence model performance. For BERT1 and BERT2, we arrived at an optimal learning rate of 0.0001, while RoBERTa's configuration required a lower rate of 0.00002, reflecting its sensitivity to learning speed. ELECTRA, on the other hand, performed best with a slightly higher learning rate of 0.00005, possibly due to its distinctive pre-training approach. All models were trained using a consistent batch size of 128, which provided a balance between resource allocation and model update frequency. The Adam optimizer was employed across all models, known for its adaptive learning rate capabilities, contributing to the fine-tuning process's efficiency and effectiveness.

Results and Observations

Streamlit UI and Example Execution

Figure 1: Stream Lit User Input Tab

The above figure allows users to input number of comments, posts, subreddit name, time filer, and interval to retrieving live reddit posts and comments.
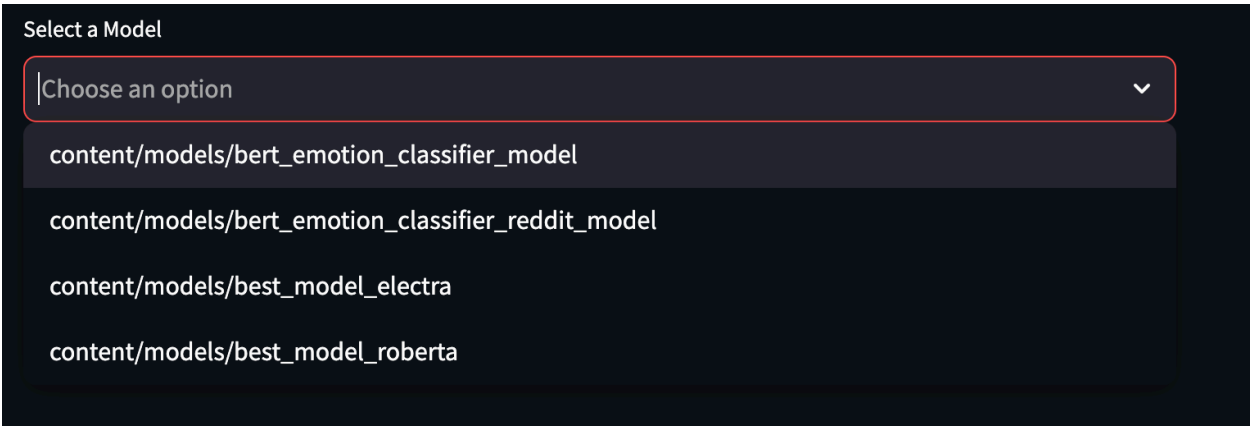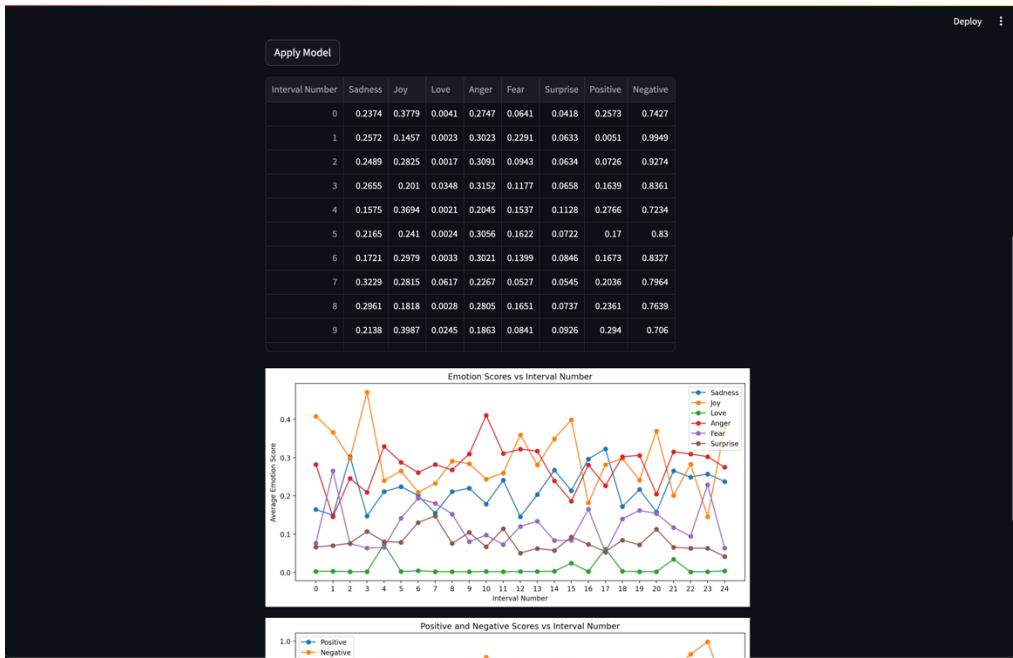

Figure 2: Transformer model Selection Tab.

Figure 2: Selecting best_model_electra

The above figure represents the output after selecting the best_model_electra and applying the fetched live reddit data on it.
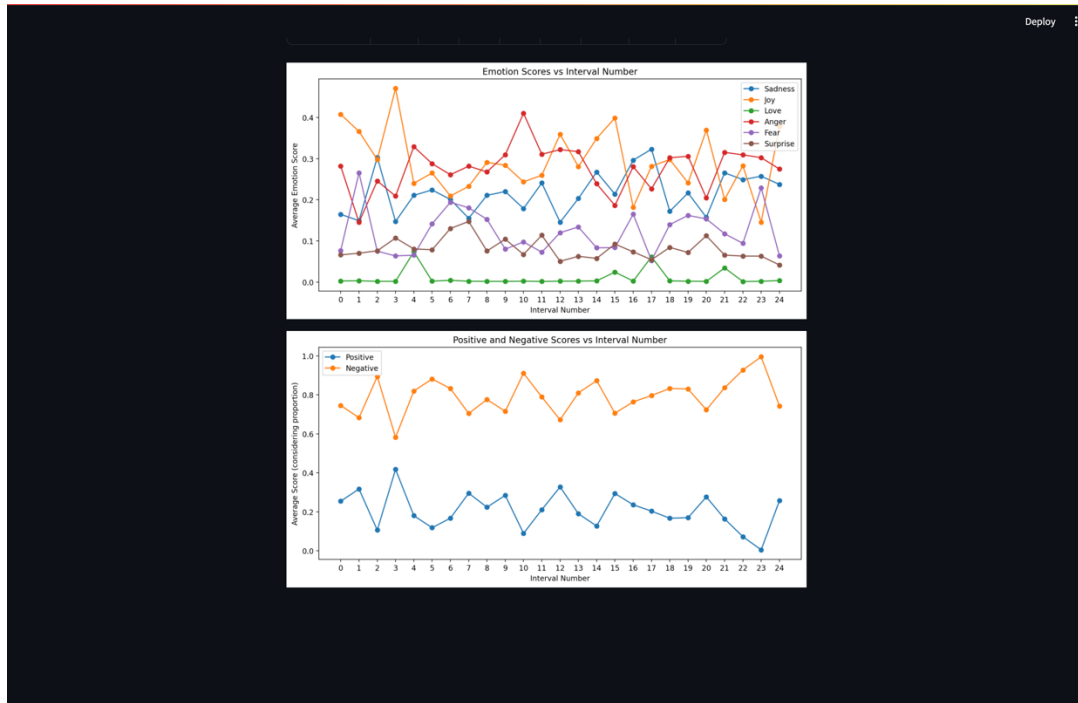
Figure 2: Final plots for sentiment trends

The first graph on the top shows the trend of average scores for each interval over a period specified by the user. Similarly, the second graph shows the average positive and negative sentiment trends to help correlate the positivity and negativity with the emotions trend.

We have also implemented the feature to use pre-loaded data to plot the sentiment trends, which can be seen from the below figures.

Figure 3: Sample execution using Pre-loaded Data (1)

| Interval Number	Sadness	Joy	Love	Anger	Fear	Surprise	Positive	Negative
0	0.2877	0.3272	0.0027	0.1329	0.1562	0.0933	0.2863	0.7137
1	0.3298	0.308	0.0017	0.2437	0.0143	0.1026	0.0098	0.9902
2	0.2522	0.4263	0.0019	0.1952	0.021	0.1035	0.2542	0.7458
3	0.2077	0.2651	0.0024	0.2124	0.2049	0.1075	0.2706	0.7294
4	0.1677	0.2988	0.0055	0.2887	0.1418	0.0974	0.2424	0.7576
5	0.2578	0.3354	0.0083	0.2705	0.0423	0.0858	0.1345	0.8655
6	0.2772	0.1713	0.0026	0.3576	0.0869	0.1045	0.2055	0.7945
7	0.2016	0.306	0.0037	0.14	0.232	0.1167	0.2138	0.7862
8	0.346	0.2592	0.0018	0.2276	0.0721	0.0932	0.2014	0.7986
9	0.183	0.3635	0.0031	0.1627	0.1862	0.1015	0.4432	0.5568
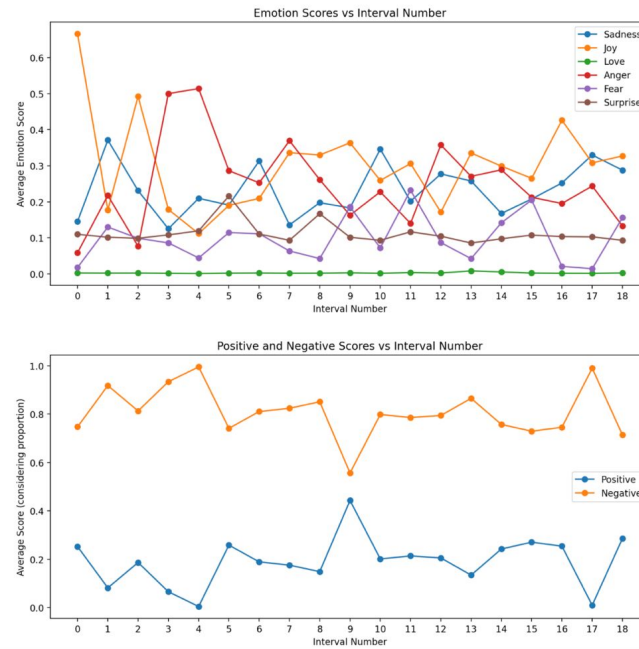
Figure 4: Sample execution using Pre-loaded Data (2)

All our models identified and predicted similar trends in the changes of sentiment, i.e., they each displayed similar deltas in emotion across time. However, while the BERT-based architectures provided robust accuracy, qualitative analysis showed ELECTRA better captured contextual emotion semantics. This was evident when predicting sentiment across different subreddits using all four of our models and comparing their visualizations. Examples with the "funny" and "wallstreetbets" subreddits show how each model identifies similar patterns, and how the ELECTRA better captures emotions that would be expected from the subreddits.
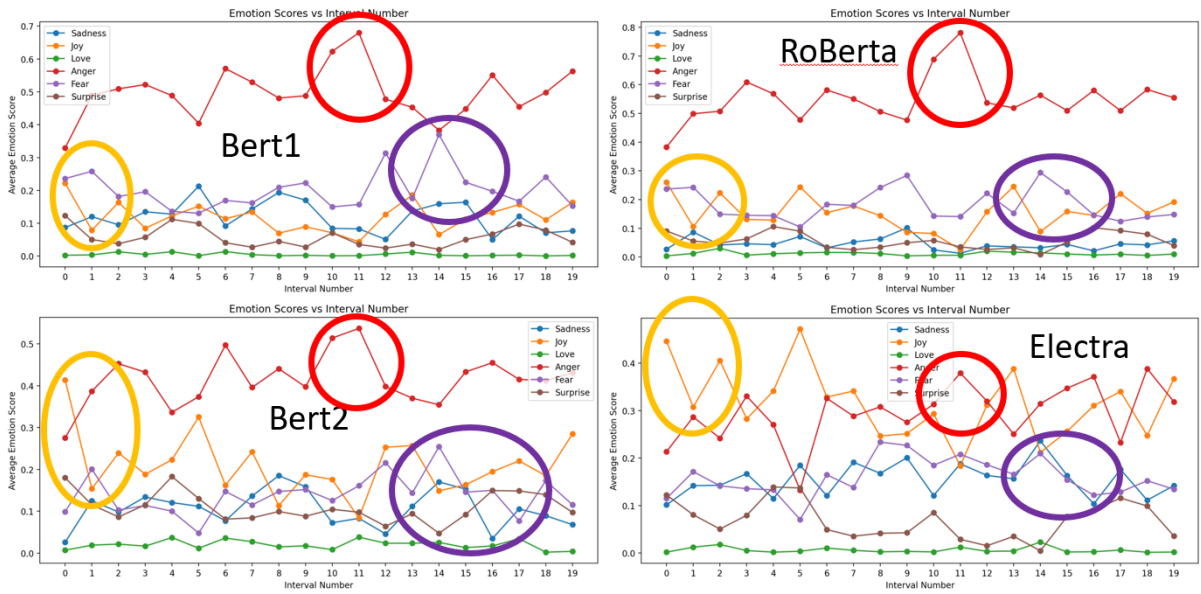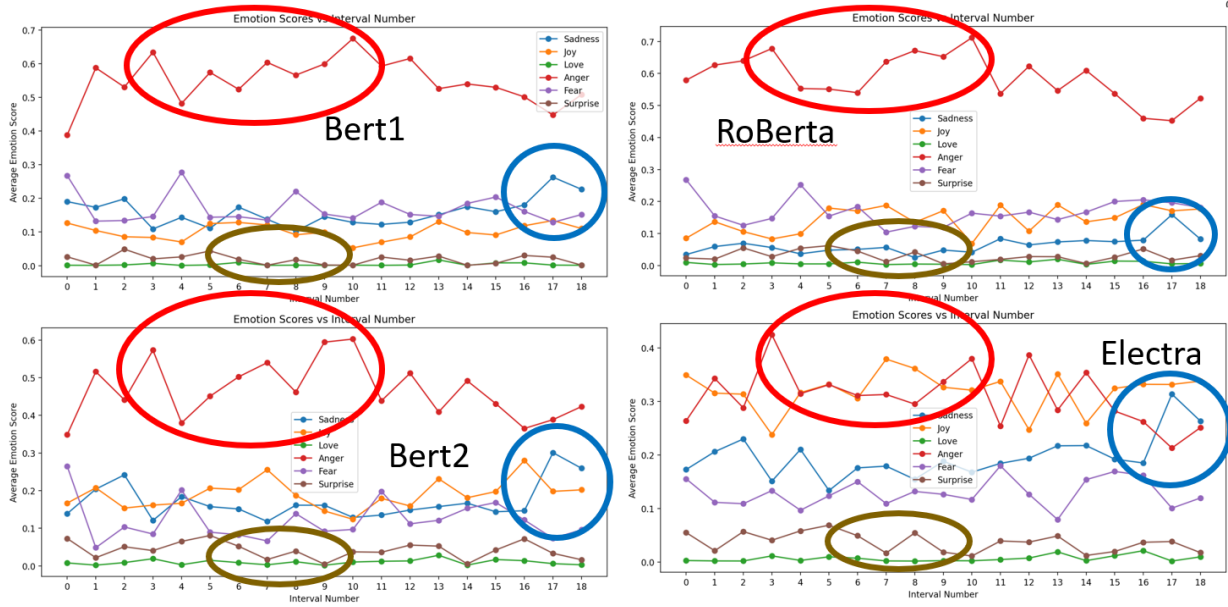
Figure 5: Qualitative Analysis of r/funny

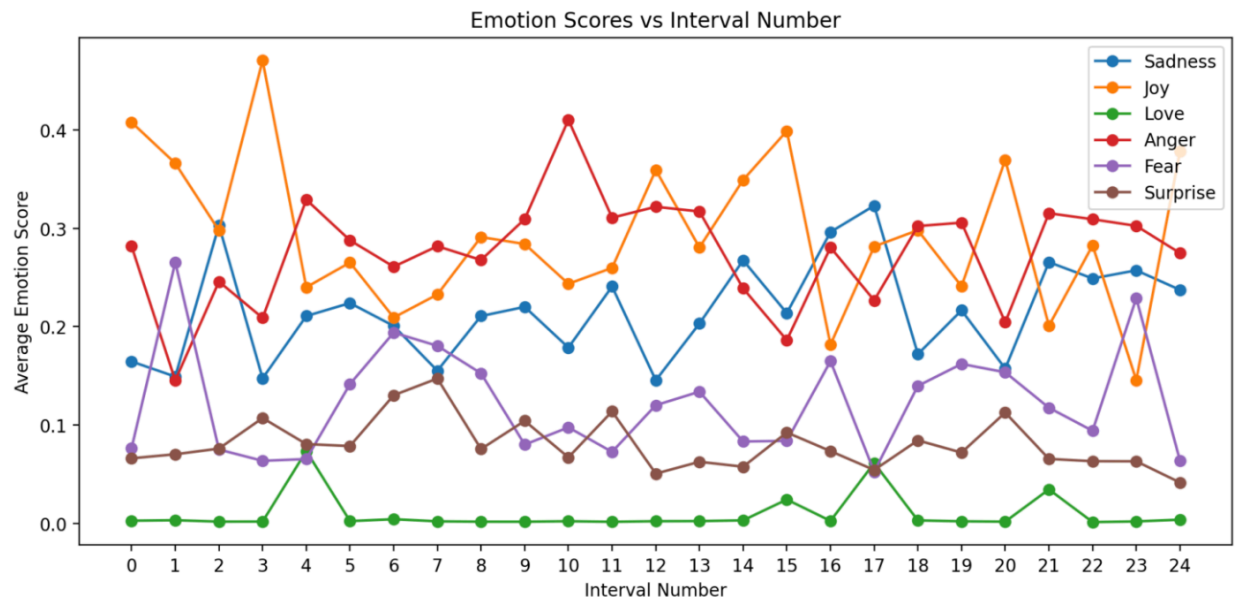Figure 6: Qualitative Analysis of r/wallstreetbets

Figure 7: Average Sentiment Score Plot of r/wallstreetbets

The figure above showcases a line graph depicting the average sentiment scores for six emotions across different intervals. A higher interval number corresponds to more recent data, while lower interval numbers indicate older data. From the graph, we can observe that members of r/wallstreetbets predominantly exhibit emotions of Anger, potentially due to financial losses in stocks, or Joy, likely resulting from profits. These predominant emotions are followed by Sadness, Fear, and Surprise, which are common in the context of r/wallstreetbets. The emotion of Love is observed the least as expected. As we can see the sentiment trend is characteristic of the r/wallstreetbets community.

Summary and Conclusions

Our project involved developing a sentiment analysis tool using advanced NLP models—BERT1, BERT2, RoBERTa, and ELECTRA—fine-tuned on a dataset of 90,000 Twitter posts labeled with six emotions and Reddit comments from r/wallstreetbets. We optimized the models by conducting a hyperparameter search, particularly focusing on learning rates and batch size, and utilized the Adam optimizer for training efficiency. The models' performance was quantitatively assessed using F1 scores and accuracy, and qualitatively by examining sentiment trends against real-world events. We integrated these models into a Streamlit application that visually displays sentiment trends over time, allowing users to interact with and analyze the changing online discourse within specified intervals.

Future Improvements

Custom Date Range Retrieval: Modify the Reddit scraping script to enable custom date range inputs for more personalized data extraction.

Dynamic Sentiment Analysis: Establish an automated system that continuously integrates, evaluates, and forecasts new data to provide up-to-the-minute sentiment analysis.

Enhanced Token Utilization: Expand the max_len parameter in pre-trained models from Hugging Face to include a greater number of tokens, enhancing the model's ability to discern subtleties in lengthier text segments, which could lead to more accurate sentiment interpretation.

Data Augmentation and Pseudo-Labeling: Enrich the training dataset by collecting additional data and employing pseudo-labeling techniques to expand the model's learning scope with unsupervised data.

References

[1] https://praw.readthedocs.io/en/stable/, PRAW (Python Reddit API Wrapper) Documentation.

[2] https://huggingface.co/docs/transformers/model_doc/electra, ELECTRA Documentation.

[3] https://huggingface.co/docs/transformers/model_doc/roberta, RoBERTa Documentation.

[4] https://matplotlib.org/stable/users/explain/quick_start.html, Matplotlib Documentation.

[5] https://towardsdatascience.com/understanding-electra-and-training-an-electra-language-model-3d33e3a9660d, Understanding ELECTRA and Training an ELECTRA Language Model.

[6] https://docs.streamlit.io/, Streamlit Documentation.

[7] https://huggingface.co/datasets/philschmid/emotion/tree/main/data, Twitter Dataset.

[8] https://www.kaggle.com/datasets/prakharrathi25/reddit-data-huge/data, Reddit Dataset.