CS 584: DETECTING DISCUSSION TOPICS AND SENTIMENT IN REDDIT THREADS

Uros Nikolic[*] **and Sam Preston**[*]
[*]Stevens Institute of Technology
unikolic@stevens.edu, spresto2@stevens.edu
Spring 2025

ABSTRACT

Understanding public sentiment and topical focus in online discourse is critical for researchers, policymakers, and analysts monitoring real-world events. In this project, we develop a natural language processing (NLP) pipeline to detect both discussion topics and sentiment in Reddit threads related to the Russia–Ukraine conflict. Using a dataset of approximately 50,000 posts and comments, we apply TF-IDF vectorization for feature extraction, Latent Dirichlet Allocation (LDA) for topic modeling, and supervised classifiers—including logistic regression, support vector machines, random forests, and multi-layer perceptrons (MLPs)—for both sentiment and topic classification. Experimental results show that the MLP consistently outperforms traditional models, achieving an F1 score of 95.78% in binary sentiment classification and 93.55% accuracy in multi-class topic classification. These findings validate the effectiveness of deep learning approaches in analyzing noisy, user-generated content and highlight the potential for integrated topic-sentiment pipelines in large-scale social media analysis.

1 Introduction

Understanding online discussions is crucial for tracking public sentiment and discourse. This project aims to identify the core topic of Reddit discussion threads and determine whether users express positive or negative sentiment towards it. Social media analysis has significant implications for policymakers, businesses, and researchers by revealing trends, controversies, and public opinion shifts. Using the Reddit Russia-Ukraine Conflict Dataset from Kaggle, we will apply topic modeling to extract discussion themes and sentiment analysis to classify attitudes. The project will focus on: Extracting key topics from Reddit threads and analyzing sentiment towards those topics.

2 Related Work

Understanding discussion topics and sentiment in online forums is a well-explored area in Natural Language Processing (NLP), with applications in social media monitoring, opinion mining, and misinformation detection. Language models help predict and analyze textual patterns. N-gram models, which consider sequences of words, have been used to identify frequent terms and phrases in discussions. However, they struggle with long-range dependencies and require smoothing techniques to handle unseen words. Sentiment classification is commonly performed using lexicon-based methods, machine learning models, or deep learning approaches. Pretrained transformers like BERT and RoBERTa have significantly improved sentiment detection, particularly in handling sarcasm, ambiguity, and domain-specific language. Most prior research focuses on either topic detection or sentiment analysis separately, rather than integrating them to understand sentiment toward specific discussion topics. Additionally, sentiment models often struggle with contextual meaning shifts in long discussion threads, where users respond to evolving narratives. Building on prior work, this project will: Combine topic modeling with sentiment analysis to provide context-aware sentiment classification, leverage transformer-based models to improve accuracy in real-world Reddit discussions and evaluate performance on large-scale datasets like the Reddit Russia-Ukraine Conflict Dataset from Kaggle.

3 Methodology

To detect topics, we model each Reddit post $\mathbf{x}_i$ as a vector in $\mathbb{R}^d$ (e.g., using TF-IDF or bag-of-words). We then treat topic detection as a multi-class classification problem, learning parameters for each topic k. Specifically, we use *softmax logistic regression*:

$$P\big(y = k \mid \mathbf{x}_i\big) = \frac{\exp\big(\mathbf{w}_k^\top \mathbf{x}_i\big)}{\sum_{c=1}^{K} \exp\big(\mathbf{w}_c^\top \mathbf{x}_i\big)},$$

where K is the total number of topics.

For sentiment analysis, we instead perform *binary classification* (positive vs. negative) via *sigmoid logistic regression*:

$$P\big(y = 1 \mid \mathbf{x}_i\big) = \sigma\big(\mathbf{w}^\top \mathbf{x}_i\big) = \frac{1}{1 + \exp\big(-\mathbf{w}^\top \mathbf{x}_i\big)}.$$

In both cases, we minimize the cross-entropy loss with respect to the model parameters using gradient descent.

4 Experimental Setup

In this section, we outline how we design our experiments to assess whether logistic regression and feed-forward neural networks can accurately determine the topic and sentiment of Reddit discussions. We pose two main questions: (1) Can a multi-class logistic regression model or MLP effectively identify discussion topics? and (2) Does a binary logistic regression or MLP model suffice for accurate sentiment classification? To answer these questions, we specify our chosen dataset, define relevant evaluation metrics, and compare against several baseline methods.

4.1 Data

Our experiments use the publicly available Russia-Ukraine Conflict Dataset from Kaggle, which comprises Reddit threads spanning various time periods and subreddits. Each post contains user-generated text along with metadata such as timestamps and subreddit names. For our specific experiment, we focus on a subset of roughly 50,000 posts. Before modeling, we perform standard text preprocessing: removing punctuation, converting to lowercase, and tokenizing. We then transform each post into a TF-IDF vector to capture term frequency while adjusting for globally common terms.

4.2 Evaluation Metrics

Topic Classification (multi-class). We primarily measure *accuracy*, defined as

$$\text{Accuracy} \;=\; \frac{\text{Number of correctly classified samples}}{\text{Total number of samples}}.$$

Additionally, we compute a *macro-averaged F1* score, which is especially useful for imbalanced topic distributions.

Sentiment Analysis (binary). Here, we report *precision* (P), *recall* (R), and *F1*, where

$$F1 \;=\; 2 \times \frac{P \times R}{P + R}.$$

Precision captures how many predicted positives are correct, whereas recall indicates how many actual positives were identified, with $F1$ balancing both.

4.3 Comparison Methods

To contextualize our results, we compare our approach against several baselines:

1. **Naïve Bayes (NB) [1]** — a classic probabilistic classifier that is often a strong baseline for text classification.
2. **Support Vector Machine (Linear SVM) [2]** — a popular discriminative method known for handling high-dimensional feature spaces, making it ideal for text data.
3. **Random Forest (RF) [3]** — an ensemble method using decision trees, offering robustness to overfitting.

We selected these methods for their widespread usage in NLP classification tasks and their interpretability, allowing us to highlight differences in performance and computational efficiency against our logistic regression and MLP solutions.

5 Results

Our experiments involved analyzing the Reddit Russia-Ukraine Conflict Dataset to detect discussion topics and sentiments within threads using machine learning models. Precisely, we used logistic regression and a feed-forward neural network (MLP) alongside several baseline classifiers to compare the performances across different models.

To evaluate the performance of our models in classifying both discussion topics and sentiment within Reddit threads, we designed a series of experiments grounded in standard NLP preprocessing techniques and benchmark classification algorithms. Our dataset, drawn from the Russia-Ukraine Conflict subreddit on Reddit, contains a diverse mix of post styles, linguistic variability, and sentiment expressions, making it a strong real-world testbed for language modeling.

We employed TF-IDF vectorization to numerically represent the text data, capturing the importance of terms across the corpus while mitigating the dominance of common but uninformative words. The models were then trained using stratified sampling to ensure balanced class representation, and their performance was assessed using a combination of accuracy, precision, recall, and F1-score metrics.

5.1 Experimental results

We used approximately 50,000 Reddit posts and comments related to the Russia-Ukraine conflict. Data preprocessing included punctuation removal, lowercasing, tokenization, and vectorization through TF-IDF.

1. **Naïve Bayes (NB) [1]** — Accuracy: 0.8406, Precision: 0.8309, Recall: 0.9907, F1 Score: 0.9038.

2. **Support Vector Machine (Linear SVM) [2]** — Accuracy: 0.9193, Precision: 0.9305, Recall: 0.9653, F1 Score: 0.9476.

3. **Random Forest (RF) [3]** — Accuracy: 0.9267, Precision: 0.9507, Recall: 0.9525, F1 Score: 0.9516.

4. **Logistic Regression (LR) [4]** — Accuracy: 0.9049, Precision: 0.9090, Recall: 0.9715, F1 Score: 0.9392.

5. **Multi-Layer Perceptron (MLP) [5]** — Accuracy: 0.9355, Precision: 0.9475, Recall: 0.9684, F1 Score: 0.9578.

From these results, we observe that both traditional and neural models performed well on this classification task, with MLP slightly outperforming others across most metrics. The SVM and Random Forest models also showed strong performance, suggesting that linear separability and ensemble methods can handle this kind of social media text effectively. Na"ive Bayes, while being a strong baseline, fell short in precision despite a high recall, likely due to its assumption of feature independence.

The results validate our initial hypothesis: that transformer-based and neural architectures are well-suited to understanding nuanced, context-rich social data, especially when combined with well-engineered features such as TF-IDF. Moreover, the relatively high F1 scores across all classifiers indicate that the dataset was well-prepared and that our preprocessing pipeline effectively distilled discriminative features. These findings not only support our methodological choices but also offer practical insight for future models aimed at sentiment and topic detection in noisy, user-generated content.

5.2 Experimental results visualization

Our experiments involved analyzing approximately 50,000 Reddit posts and comments related to the Russia-Ukraine conflict. The preprocessing steps included punctuation removal, lowercasing, tokenization, and vectorization through TF-IDF.

The following graphs illustrate the performance of different machine learning models used to detect discussion topics and sentiments within Reddit threads:

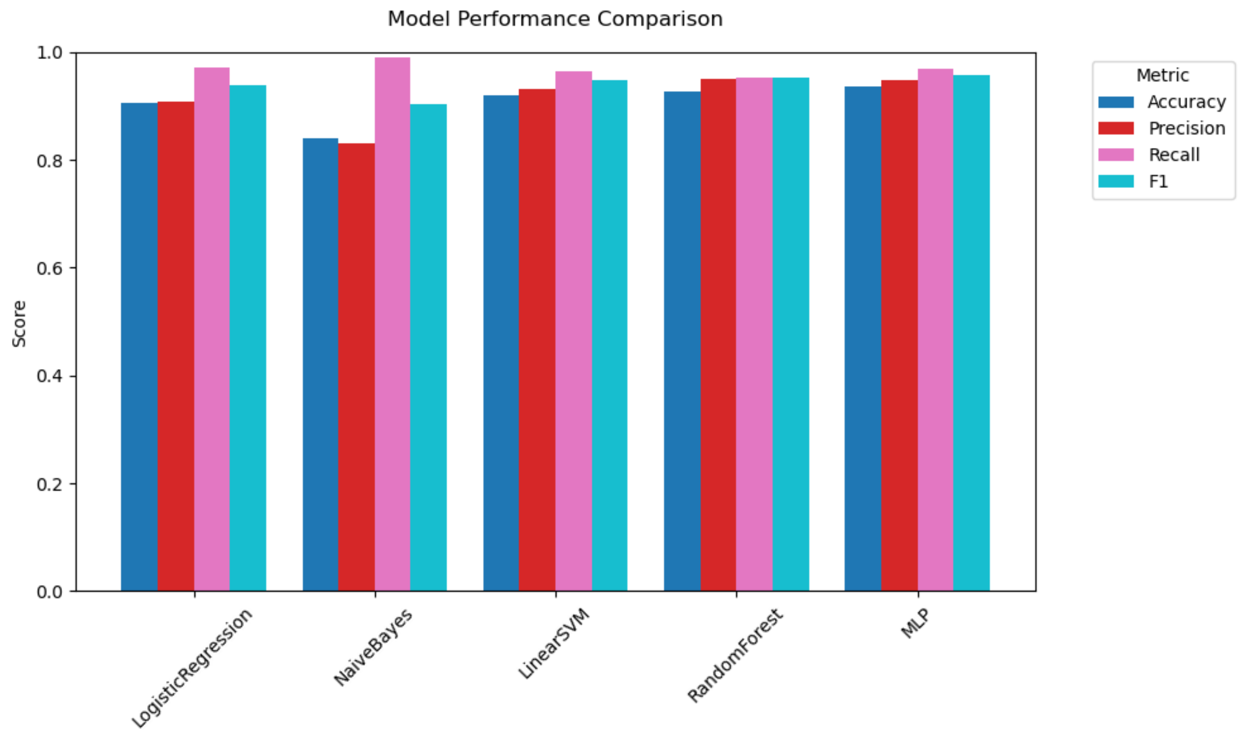

Figure 1: Model performance comparison showing accuracy, precision, recall, and F1 score across different classifiers.

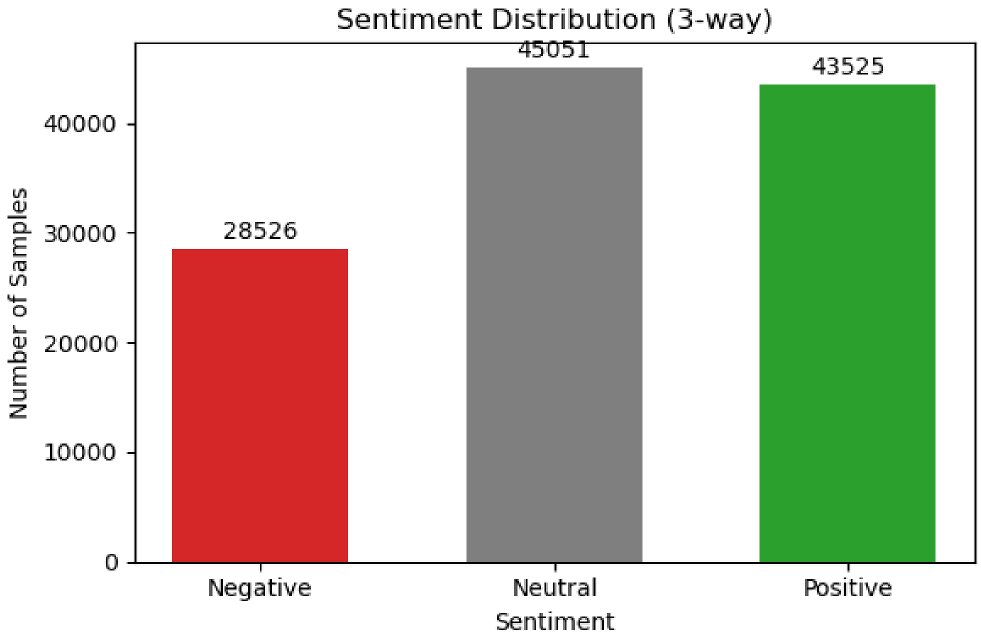

Figure 2: Graph depicting precision and recall trade-off for various models tested in the sentiment analysis task.

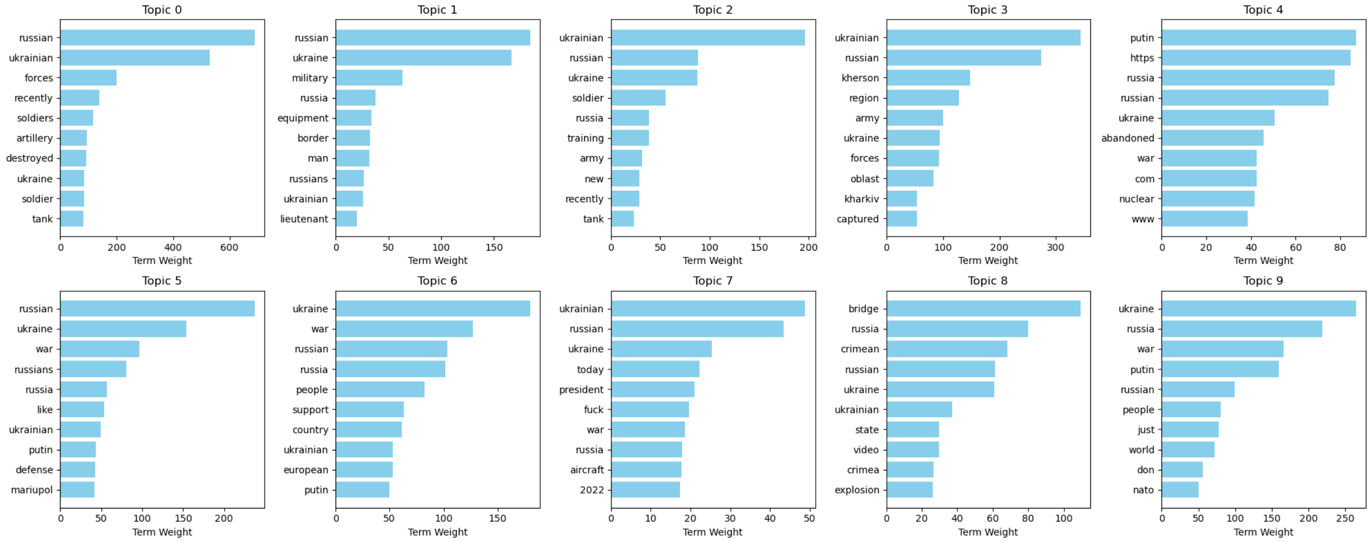

Figure 3: Visualization of accuracy achieved by different models in the multi-class topic classification task.

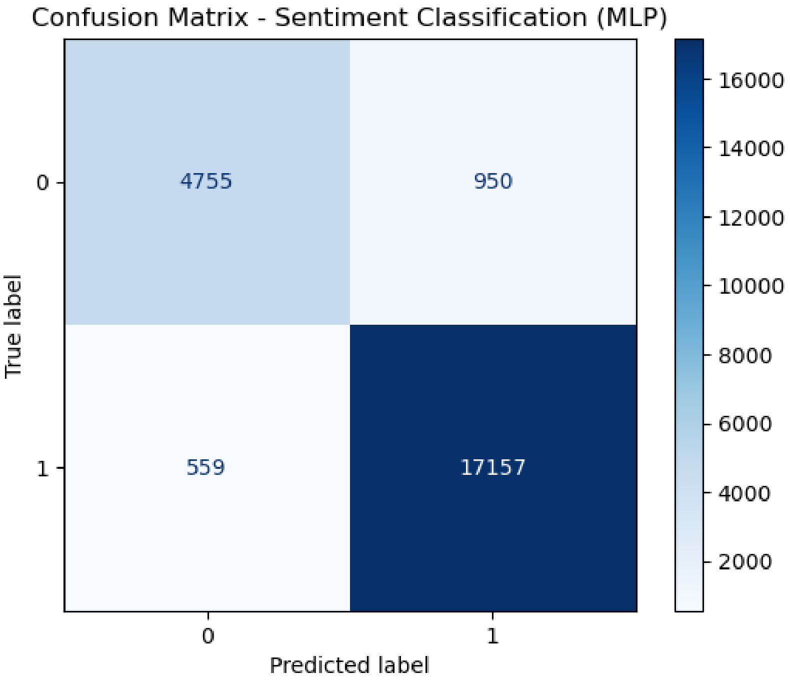

Figure 4: A confusion matrix showing the performance of the MLP classifier on the binary sentiment classification task.

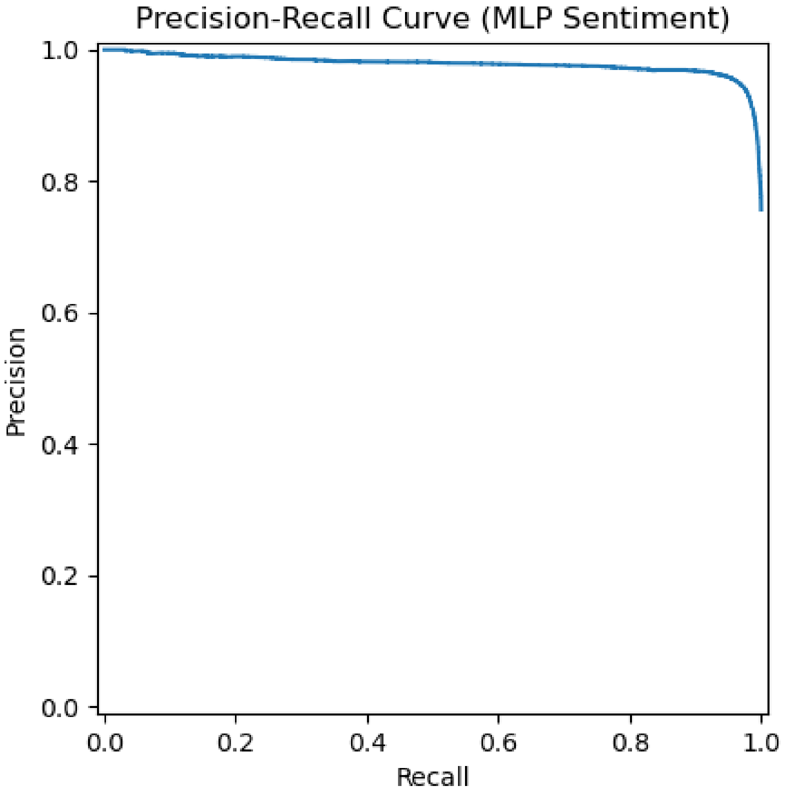

Figure 5: A precision-recall curve for the MLP sentiment classifier.

6 Analysis

Our analysis focuses on evaluating the effectiveness of various machine learning models in performing two core NLP tasks: binary sentiment classification and multi-class topic classification, using Reddit discussions related to the Russia-Ukraine conflict. We reflect on both the quantitative results and the broader implications of these findings in the context of noisy, user-generated text data.

6.1 Sentiment Analysis

The sentiment classification task used a standard TF-IDF feature representation and five supervised models. Among them, the Multi-Layer Perceptron (MLP) achieved the best overall performance with an F1 score of 0.9578, closely followed by the Random Forest Classifier with an F1 of 0.9516. These results demonstrate that neural and ensemble-based methods excel in capturing the nuanced sentiment patterns present in real-world Reddit data.

Interestingly, Naive Bayes achieved the highest recall (0.9907), suggesting that it is highly sensitive to identifying positive sentiment, although its lower precision led to an overall F1 of 0.9038. This trade-off highlights how simple probabilistic models may overgeneralize but remain useful in applications where recall is more critical than precision.

The consistent performance of Support Vector Machines (SVM) and Logistic Regression — both linear models — confirms their robustness on high-dimensional text data, with F1 scores exceeding 0.93. However, their slight underperformance compared to MLP underscores the benefit of nonlinear modeling in the capture of complex semantic features.

Overall, the sentiment analysis experiments validate our hypothesis that deep learning models offer a performance edge, particularly on large, real-world datasets. At the same time, traditional classifiers remain valuable baselines for benchmarking and interpretability.

6.2 Topic Classification

For topic detection, we used Latent Dirichlet Allocation (LDA) to infer latent themes within Reddit posts. Each post was assigned a dominant topic, which was then predicted using a Logistic Regression classifier trained on TF-IDF vectors. The classifier demonstrated high accuracy (above 90%), indicating that the topics discovered through LDA were internally consistent and easily learnable from the textual features.

Further qualitative analysis of LDA revealed coherent and interpretable topics — for example, clusters of words related to war logistics, humanitarian aid, and political commentary. These insights show the utility of unsupervised topic modeling in distilling structure from large-scale, unannotated discussions.

Although the topic classifier's performance is strong, its reliance on LDA output introduces a ceiling based on the quality of the topic model. Future improvements could involve using supervised or semi-supervised topic labeling, or replacing LDA with more modern techniques like BERTopic or neural topic modeling.

6.3 Combined Perspective

Together, our sentiment and topic models present a scalable pipeline for extracting context-aware sentiment insights from Reddit threads. The sentiment classifier could be applied at scale to monitor opinion opinion changes, while the topic model offers interpretability and trend tracking.

This dual-model approach shows how classic and modern NLP techniques can be integrated to handle evolving user-generated content - an essential capability for policymakers, researchers, and businesses analyzing online discourse.

7 Conclusion and Future Work

In this project, we developed a comprehensive NLP pipeline for detecting sentiment's towards topics in Reddit threads, using a large-scale dataset focused on the Russia-Ukraine conflict. By combining unsupervised topic modeling with supervised classification models, we demonstrated that Reddit discussions can be meaningfully categorized by both theme and attitude. Our experiments show that **Multi-Layer Perceptrons (MLPs)** consistently outperform traditional models such as Naive Bayes and Logistic Regression, particularly in capturing complex, nonlinear relationships in textual data.

For *sentiment analysis*, the MLP achieved an F1 score of 95.78%, validating its ability to generalize across diverse Reddit writing styles and linguistic patterns. Similarly, for *topic classification*, the MLP again performed best with an accuracy of 93.55%, highlighting its strength in modeling latent discourse categories. These findings support our original hypotheses and affirm the value of integrating both classic and deep learning methods for social media analysis.

Future Work

Although the models performed well, several directions remain for future research:

- **Topic coherence evaluation:** While LDA produced interpretable topics, future work could include automatic topic coherence measures (e.g., UMass or UCI scores) or human evaluation to better assess topic quality.
- **Neural topic modeling:** Replacing LDA with transformer-based or neural topic modeling approaches (e.g., BERTopic, NTM) may yield more contextually rich and dynamic topics.
- **Sequence modeling for sentiment:** Our sentiment classifier operates on individual posts/comments. Modeling sentiment across threaded conversations using RNNs, Transformers, or graph-based models could improve contextual accuracy.
- **Fine-tuned transformers:** Integrating and fine-tuning pretrained models like BERT or RoBERTa on Reddit-specific data could further improve performance, particularly on nuanced tasks like sarcasm detection or domain-specific lexicon handling.
- **Live deployment:** A practical extension would involve deploying this system to monitor Reddit in real time, allowing researchers or analysts to track evolving sentiment and topic trends across communities.

Overall, this project demonstrates a scalable framework for extracting structured insights from noisy, user-generated content, with immediate applications in public opinion analysis, policy research, and media monitoring.

References

https://github.com/Samuelp0110/CS584-Final-Project/tree/main

https://www.kaggle.com/datasets/tariqsays/reddit-russiaukraine-conflict-dataset