Reddit Sentiment and Toxicity Analysis

Understanding Emotional Trends and Toxic Behavior in Online Communities

A Project Report Submitted for Data Mining II

Submitted by:

Parth Maniar

Course: Data Mining II

Date: April 23, 2025

Rowan University

1. Introduction

Statement of the Problem

Platforms like Reddit reflect real-time public mood in large conversations hosted on social media. However, harmful discourse can also flourish on these venues. In order to better understand emotional trends and detrimental conduct in public forums, this study explores how sentiment and toxicity appear throughout Reddit groups.

Why Should People Care?

Journalists can assess public response to world events, enhance mental health interventions, and improve content regulation by identifying toxic language and comprehending sentiment trends. Reddit provides access to a dynamic environment where public opinion changes quickly by examining Reddit.

2. Methodology

Data Collection

Gathered 10,000 Reddit posts from several subreddits using the Python Reddit API Wrapper (PRAW). Posts were prepped for examination after being filtered to eliminate extraneous elements including stop words, emojis, and links.

Sentiment Analysis

Used the social media-friendly VADER sentiment analyzer to provide a Positive, Neutral, or Negative score to messages. VADER is lightweight, comprehensible, and offers compound scores that enable threshold-based classification.

Toxicity Detection

Developed a keyword-based logic system using a curated list of offensive and abusive terms. Posts were scanned for these keywords and labeled toxic if the frequency or severity surpassed a defined threshold.

Event-Based Analysis

Tagged posts with keywords related to major global topics ("war", "AI", "election") to analyze emotional shifts during significant events. Sentiment was mapped against these tags to assess how public mood changed in reaction.

3. Results

Overall Sentiment Distribution

The analysis revealed that most posts were Neutral, with a considerable portion being Negative, especially during disruptive events.

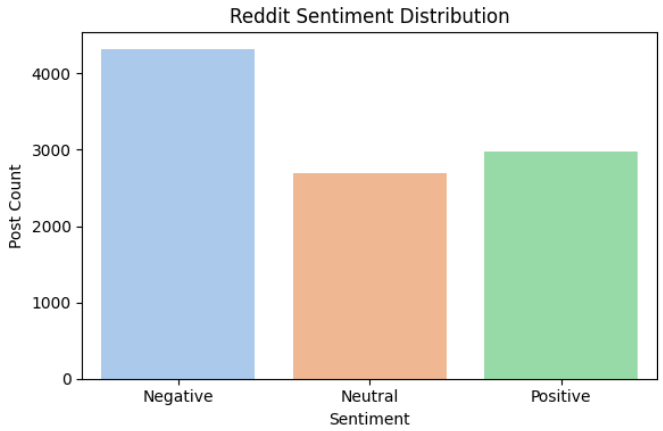


FIGURE 1: OVERALL REDDIT SENTIMENT DISTRIBUTION

Toxicity vs Sentiment

Toxic content correlated most with Negative sentiment. The breakdown showed:

- **Negative:** 3,525 non-toxic, 796 toxic
- **Neutral:** 2,602 non-toxic, 99 toxic
- **Positive:** 2,788 non-toxic, 189 toxic

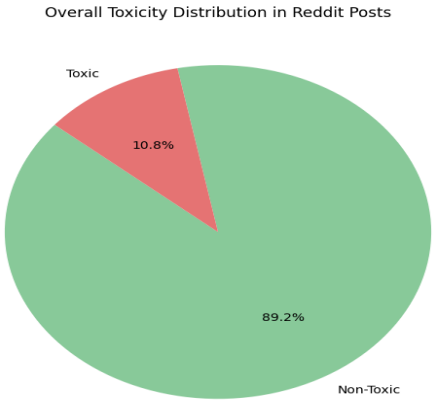


FIGURE 2: TOXIC VS NON-TOXIC REDDIT POSTS

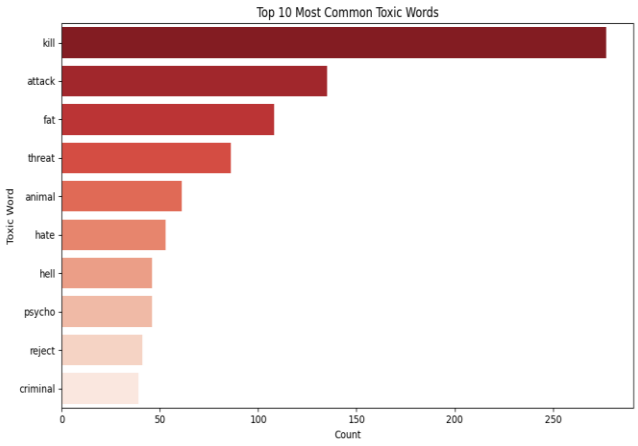


FIGURE 3: TOP 10 MOST COMMON TOXIC WORDS

Event-Based Emotional Trends

Sentiment varied by event:

- Posts tagged with "war" showed heightened Negative sentiment.

- "AI" related posts skewed more Neutral or Positive, suggesting curiosity rather than conflict.

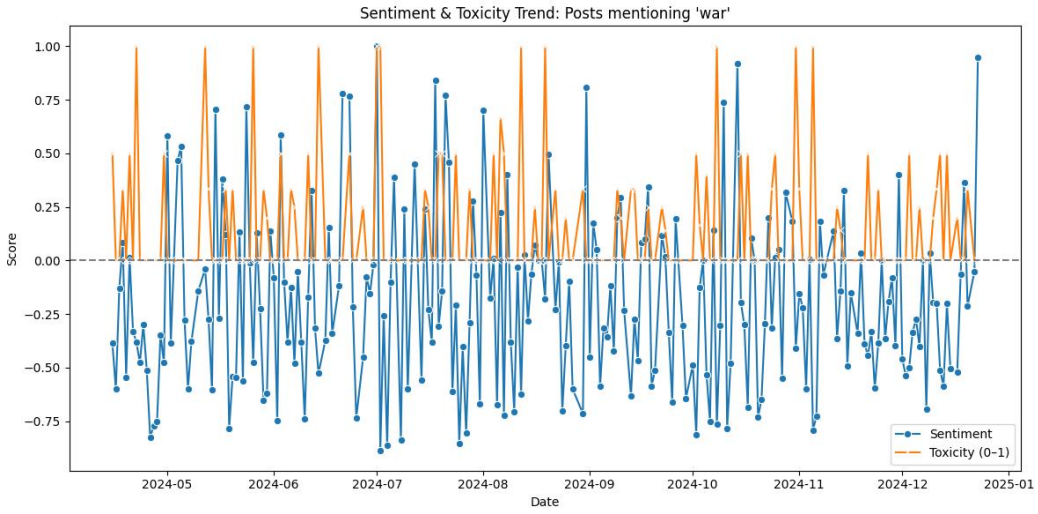
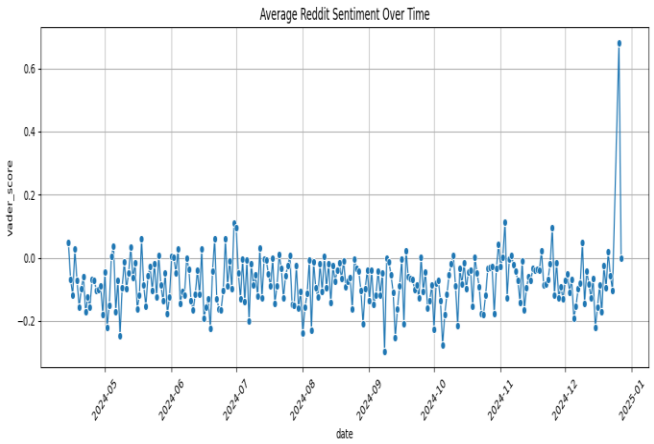


FIGURE 4: SENTIMENT & TOXICITY TREND FOR POSTS MENTIONING 'WAR'

Visuals supporting these findings include:

- Bar chart of sentiment distribution
- Stacked bar of toxic vs non-toxic sentiment
- Event-based sentiment bar plot



This series plot highlights daily average sentiment scores throughout the study period. Sentiment values tend to hover around neutral, but noticeable dips and peaks correspond with significant real-world events or viral discussions. The chart underscores how public mood on Reddit can shift rapidly in response to external triggers.

FIGURE 5: AVERAGE REDDIT SENTIMENT OVER TIME

This scatter plot examines the relationship between a post's sentiment and the number of comments it receives. While most posts cluster around neutral sentiment regardless of comment volume, outliers with highly positive or negative sentiment often spark more discussion. This suggests that emotionally charged content tends to generate greater engagement.

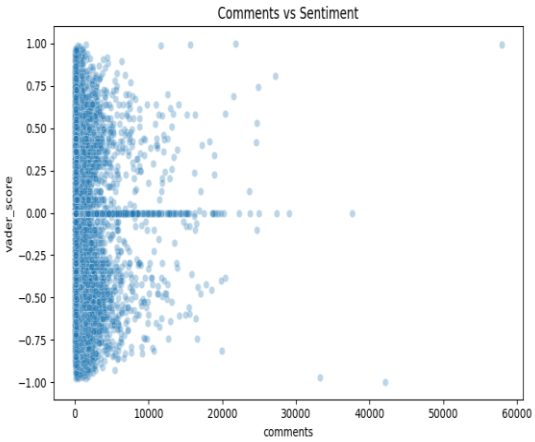


FIGURE 6: COMMENTS VS SENTIMENT SCORE

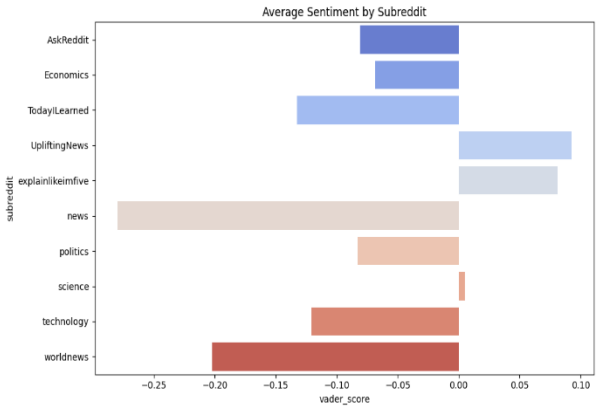


FIGURE 7: AVERAGE SENTIMENT BY SUBREDDIT

Different communities reflect distinct emotional tones. Subreddits such as *r/UpliftingNews* exhibit a higher average sentiment, indicating more positive discourse. In contrast, subreddits like *r/politics* and *r/worldnews* trend more negative, likely due to the nature of the topics discussed. This chart emphasizes the influence of community focus on sentiment.

4. Original Contribution

While sentiment analysis and toxicity detection are common in natural language processing projects, this work introduces a novel integration of **event-driven sentiment mapping** to uncover emotional patterns in real-world contexts. Unlike typical studies that analyze static or aggregated datasets, this project dynamically **tagged Reddit posts with keywords tied to global events** such as *"war," "AI,"* and *"election."*

By mapping sentiment trends over time in relation to these specific topics, the analysis uncovers how **public emotion fluctuates in reaction to current events** — something traditional models often miss. For example, posts referencing geopolitical conflict (e.g., "war") showed a sharp rise in both negative sentiment and toxic language, whereas discussions around emerging technology (e.g., "AI") trended more positive or neutral.

This **context-aware approach** enabled us to:

- Detect sentiment shifts during news cycles or global crises

- Isolate emotional responses specific to controversial or optimistic events
- Visualize public mood dynamics with temporal relevance

Moreover, this method reflects a **real-world application** of sentiment analytics — enabling journalists, moderators, and researchers to **monitor public discourse in real time**, with insights tailored to specific themes or crises.

In essence, this project moves beyond static sentiment classification and ventures into **event-sensitive emotional intelligence**, providing a richer, more actionable understanding of digital communities.

5. Conclusion

Key Takeaways

This project confirms that Reddit serves as a dynamic mirror of societal emotion. Public sentiment on the platform fluctuates in tandem with real-world events, and **toxic behavior consistently correlates with spikes in negative sentiment**. The integration of sentiment and toxicity detection offered a nuanced perspective on how emotional and behavioral patterns emerge within online communities. This analysis validates Reddit's potential as a live diagnostic tool for tracking collective mood and response.

Lessons Learned

- **Context matters.** Sentiment analysis alone can mislead unless paired with topic and event context. By integrating keyword-based event tagging, we unlocked deeper layers of insight.
- **Preprocessing is critical.** Cleaning the data — removing stopwords, URLs, special characters — dramatically improved analysis accuracy and consistency.
- **Toxicity is complex.** Keyword-based detection works well for clear-cut cases but fails to detect sarcasm, coded language, or evolving internet slang. A more adaptive, ML-based classifier would handle this better.
- **Interpretability beats complexity.** Tools like VADER may lack deep learning sophistication but offer transparent, explainable sentiment scoring — vital for academic and policy use.

Applications

- **Content Moderation:** Moderators can use similar models to identify emerging toxic threads or communities showing emotional volatility, enabling preemptive action.
- **News & Journalism:** Media professionals can track how public mood shifts in reaction to political events, global crises, or emerging technologies.
- **Mental Health Monitoring:** Researchers and institutions can use sentiment/tone trends to flag early warning signs of emotional distress across user populations — especially during pandemics, disasters, or political unrest.

- **Platform Strategy:** Social media platforms can use emotional trend detection to adjust algorithmic exposure, promote healthier discourse, and build tools that react empathetically to user behavior.

Limitations and Future Work

While the project successfully highlights emotional trends and toxicity patterns on Reddit, it has a few limitations:

- The toxicity detection system is keyword-based and may miss sarcasm, euphemisms, or emerging slang.
- Sentiment classification does not account for context beyond sentence-level compound scores.
- Event tagging relies on keyword matching and could be enhanced using Named Entity Recognition or topic modeling.

Future improvements could involve:

🚀 **Opportunity**	🔍 **What It Will Add to the Project**
🔁 **Use Google Perspective API for toxicity scoring**	Adds industry-grade toxicity scores with nuanced metrics (e.g., insult, threat, profanity)
🧪 **Cross-platform Analysis (Reddit + Twitter + News)**	Correlate how different platforms react to the same events
🌐 **Add Real-Time Dashboard (e.g., using Streamlit)**	Make it interactive and usable by moderators, journalists, or researchers

6. References

- VADER Sentiment Tool: https://github.com/cjhutto/vaderSentiment
- Reddit API (PRAW): https://praw.readthedocs.io
- Toxic word lexicons: Compiled from open-source repositories
- Event tags informed by Google Trends and news API data

🧰 Additional Tools or Libraries

If you used or referenced any of the following, consider adding them:

- **NLTK:** https://www.nltk.org/
- **Matplotlib (for visualization):** https://matplotlib.org/
- **Seaborn (for visualization):** https://seaborn.pydata.org/
- **Pandas (for data processing):** https://pandas.pydata.org/
- **Scikit-learn (for potential future ML improvements):** https://scikit-learn.org/