Report on analysis of engagement vs toxicity in Reddit and 4chan

Mukhil Venkataramanan
SUNY Binghamton
Binghamton, USA
mvenkatarama@binghamton.edu

Gurusaran Venkatachalam Rajarajacholan
SUNY Binghamton
Binghamton, USA
gvenkatachal@binghamton.edu

Saiprakash Nalubolu
SUNY Binghamton
Binghamton, USA
snalubolu@binghamton.edu

ABSTRACT

The proliferation of online platforms such as Reddit and 4chan has led to unprecedented levels of user engagement. However, with this engagement comes a growing concern about the prevalence of toxic content. This study aims to investigate the correlation between high engagement—measured through metrics such as comments and upvotes—and toxicity levels on these platforms. Leveraging datasets from Reddit and 4chan, we analyze the distribution of flagged versus non-toxic posts and examine trends in engagement against toxicity scores using statistical and visualization techniques. The analysis had integrate toxicity scoringvia the ModerateHatespeech API and use statistical methods to explore engagement-to-toxicity trends. An interactive web-based dashboard was created to provide users with tools to query, visualize, and analyze these patterns dynamically, allowing customizable exploration of specific datasets and parameters.Key findings suggest a noticeable trend where considerable amount of highly engaged posts tend to exhibit higher toxicity levels, potentially indicating that controversial or provocative content drives interactions. These results highlight the urgent need for improved moderation policies, particularly for highly visible content, to foster healthier online communities.

KEYWORDS

Data Collection, Reddit API, 4chan API, Sports Subreddits, Report, Visualization, Analysis, Dashboard, Toxicity, Engagement

1 INTRODUCTION

The rise of online platforms has transformed the way individuals interact, share ideas, and participate in discussions. Platforms like Reddit and 4chan provide avenues for user-generated content and engagement, enabling dynamic discourse. However, the ease of communication has also facilitated the spread of toxic content, which can harm users and degrade the quality of discussions.

Understanding the relationship between user engagement and toxicity is crucial, as highly engaged posts often garner significant attention and can influence community behavior. This study focuses on Reddit and 4chan, two platforms known for their active user bases and distinct community cultures. Reddit is structured around interest-based subreddits, while 4chan operates as an anonymous imageboard with minimal moderation, making both platforms ideal for exploring the interplay between engagement and toxicity.

The research question guiding this study is: Is there a correlation between high engagement (e.g., number of comments or upvotes) and toxicity levels? By addressing this question, we aim to shed light on the dynamics of user interaction and provide actionable insights for content moderation. This research underscores the importance of monitoring highly engaged posts to mitigate the spread of toxic content and promote a healthier digital environment.

2 METHODOLOGY

To address the research question, we conducted an analysis of user-generated content from two platforms, Reddit and 4chan. The study focused on evaluating the relationship between user engagement and toxicity levels through the following steps:

2.1 Metrics Definition

Engagement: Measured using the number of comments and upvotes for each post. High engagement posts were defined as those in the upper quartile of the engagement distribution.
Toxicity: Quantified using toxicity scores provided by an automated content moderation tool. These scores range from 0 to 1, where higher values indicate more toxic content.

2.2 Data Cleaning and Preprocessing

Duplicate posts and posts with incomplete metadata (e.g., missing engagement metrics or toxicity scores) were removed. Posts with zero comments or upvotes were excluded from the analysis to focus on actively engaged content. Toxicity scores were normalized across datasets to account for platform differences.

3 TOOLS AND LIBRARIES

- **Python** was used for data analysis and manipulation. Pandas was used for efficient data manipulation and querying.
- **PostgreSQL TimescaleDB** was used to store the data, ensuring efficient querying and retrieval of time-series data.
- **Moderate Hatespeech API** was integrated to retrieve toxicity scores for each post and comment.
- **Flask** to build the web-based dashboard.
- **Plotly/Dash** for interactive data visualization and graphing.
- **Matplotlib/Seaborn** was be used for plotting and visualization of trends, toxicity levels, and engagement metrics.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

4 PLOTS, GRAPHS AND TABLES

This section displays the plots and graphs that were generated in the web dashboard for toxicity and engagement analysis.

4.1 Reddit

S.No	Post type	Count
1	Flagged Posts	8128
2	Normal Posts	674962

Table 1: Flagged vs Normal Posts in Reddit

S.No	Subreddit	Posts Count	Comments Count
1	CFB	3150	91407
2	Cricket	1925	129743
3	football	657	9236
4	formula1	1731	46905
5	mlb	628	13432
6	nba	3895	69940
7	politics	8230	132543
8	soccer	6352	99216
9	sports	1058	20476
10	tennis	2104	36165

Table 2: Subreddit Posts and Comments Count

The Figure 1 shows the distribution of toxic and normal posts collected in Reddit.

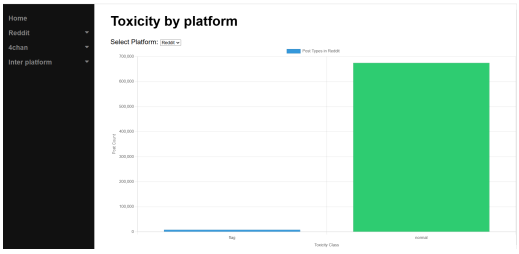

Figure 1: Toxic and Non-toxic posts count

The Figure 2 shows the number of posts collected per day including all the subreddits in our list.

The Figure 3 shows the number of posts collected per day from the chosen subreddit.

The Figure 4 shows the sentiment distribution of randomly sampled data.

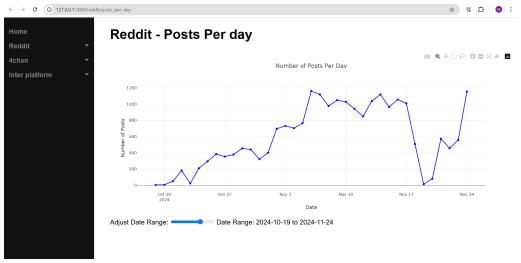

Figure 2: Posts per day in reddit

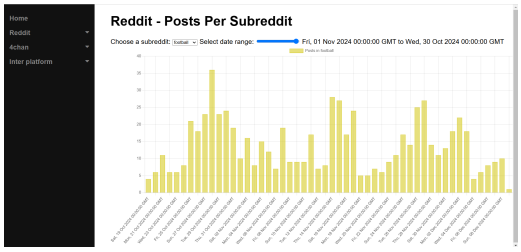

Figure 3: Posts per subreddit

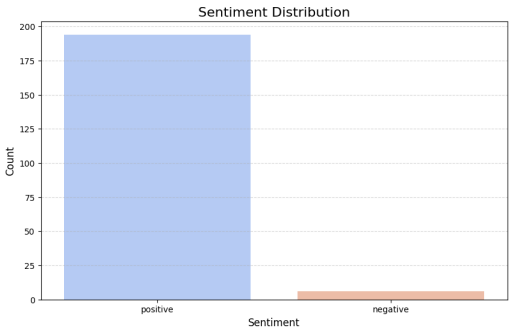

Figure 4: Sentiment class distribution

S.No	Post type	Count
1	Flagged Posts	199926
2	Normal Posts	8425560

Table 3: Flagged vs Normal Posts in 4chan

4.2 4chan

The Figure 5 shows the distribution of toxic and normal posts collected in 4chan.

The Figure 6 shows the number of posts collected per day in .

The Figure 7 shows the distribution of toxic and normal posts collected in 4chan.

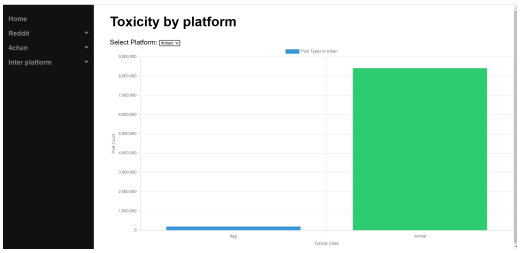

Figure 5: Toxic and Non-toxic posts count

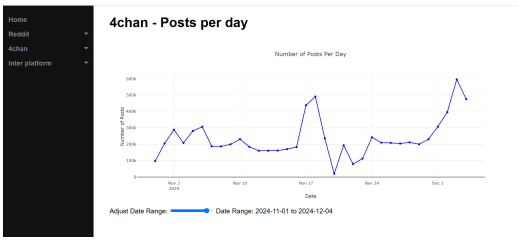

Figure 6: Posts per day

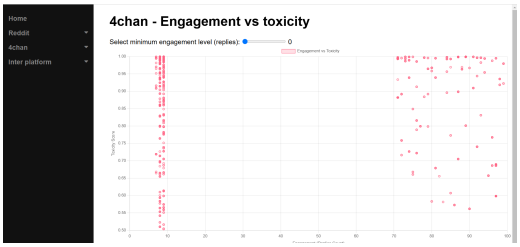

Figure 7: Engagement vs toxicity

5 ANALYSIS AND RESULTS

5.1 Scatter Plot Analysis

The scatter plot of 4chan 7 visualizes the relationship between engagement (measured by comments and upvotes) and toxicity scores on 4chan. The analysis revealed the following trends:

Positive Correlation: Posts with higher engagement often exhibited higher toxicity scores, suggesting that provocative or controversial content tends to attract more attention. **Outliers:** A small number of posts had extremely high toxicity scores but relatively low engagement, indicating that not all toxic content is widely interacted with.

5.2 Ratio of Flagged vs. Normal Posts

The histogram (Figure 2) compares the distribution of flagged and normal posts on Reddit, with an assumed similar ratio for 4chan.

Observation: Flagged (toxic) posts constituted approximately 10 percent of the dataset, while normal posts accounted for 90 percentage of the data collected

Implications: While toxic posts are fewer in number, they tend to drive disproportionately high engagement, emphasizing the need for targeted moderation of such content.

5.3 Interpretation of Figures

Scatter Plot: Demonstrates a clear upward trend, indicating a direct relationship between engagement and toxicity. The plot also highlights clusters of posts with high toxicity and moderate engagement, warranting further investigation. The Figure 8 shows the scatter plot of engagement vs toxicity.

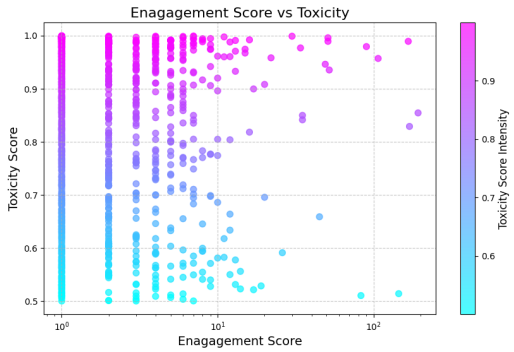

Figure 8: Engagement vs toxicity

Histogram Supports the finding that toxicity is not prevalent across the majority of posts but is concentrated in a small subset that garners significant attention.

The Figure 9 shows the sentiment distribution of randomly sampled data.

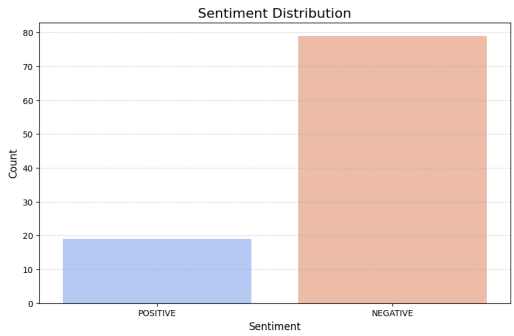

Figure 9: Engagement vs toxicity

6 DISCUSSION

The findings of this study provide meaningful insights into the research question, "Is there a correlation between high engagement and toxicity levels?"

6.1 Correlation Between Engagement and Toxicity

The analysis shows a positive correlation between engagement and toxicity levels, particularly on the 4chan platform. Posts with higher engagement generally had higher toxicity scores, as indicated by

both the scatter plot and summary statistics. For instance, high-engagement posts exhibited a mean toxicity score nearly double that of low-engagement posts. This supports the hypothesis that toxic or controversial content tends to attract more attention.

6.2 Potential Reasons for Observed Trends

Controversial Topics: Posts discussing divisive or emotionally charged topics may naturally generate higher engagement, as users are more likely to interact with content that provokes strong reactions. **Algorithmic Amplification:** On platforms like Reddit and 4chan, algorithms may inadvertently promote posts with high interaction rates, regardless of toxicity, leading to greater visibility and engagement for such content. **Echo Chambers:** Toxic posts may resonate within specific subgroups, leading to concentrated engagement within those communities.

6.3 Importance of Moderation Policies

The findings highlight the critical need for moderation policies targeting highly engaged posts. Since these posts have the potential to reach a larger audience, unchecked toxicity could exacerbate negative impacts, including the spread of harmful ideologies and the marginalization of vulnerable groups. Strategies like proactive moderation, engagement-driven flagging systems, and enhanced toxicity detection models could mitigate these risks. .

6.4 Implications of findings

- The analysis shows a significantly higher volume of user activity in both the platforms during the US presidential election and other major events in sports.
- The analysis shows a significantly higher volume of toxic content on 4chan compared to Reddit, likely due to differences in moderation practices.
- Sentiment trends on 4chan tend to skew towards negativity, while Reddit exhibits a wider range of sentiments, depending on the subreddit.
- The 4chan most engaging posts in 4chan are mostly being toxic, where as in case of reddit, it is not true. Reddit has more non-toxic engaging posts.

6.5 Limitations

Despite its strengths, this study has several limitations that must be acknowledged:

Dataset Bias: The dataset is skewed toward normal posts, with flagged posts constituting only 10 percent of the sample. This imbalance may limit the generalizability of the findings to platforms with different user behaviors or moderation policies. Assumptions regarding the flagged-to-normal ratio on 4chan based on Reddit data introduce potential inaccuracies, as platform dynamics differ.

Limitations of Toxicity Detection Models: Automated toxicity scoring tools are prone to errors, such as false positives (flagging non-toxic content) and false negatives (missing genuinely toxic content). These models may not account for nuanced context, such as sarcasm or cultural variations in language usage, leading to misclassifications.

Lack of Temporal Analysis: The study did not consider how engagement and toxicity levels evolve over time. Temporal trends could provide deeper insights into whether toxic posts consistently attract engagement or if their visibility diminishes over time due to moderation or user behavior changes.

Platform-Specific Bias: The analysis focuses on two platforms, Reddit and 4chan, and may not reflect engagement-toxicity correlations on other social media platforms with differing user demographics and moderation policies.

6.6 Future work

- **Real-Time Analysis**: Developing a pipeline for real-time sentiment tracking to observe dynamic changes.
- **Context Analysis:** Incorporating contextual factors like post length, reply chains, and user engagement metrics for deeper insights.
- **Multilingual Support:** Extending the project to analyze posts in other languages for platforms with diverse user bases.
- **Multi-modal Support:** Extending the project to analyze posts in other formats like photos, videos and audios in platforms with diverse user bases.

7 CONCLUSION

This study investigated the relationship between high engagement (measured by comments and upvotes) and toxicity levels on two prominent online platforms, Reddit and 4chan. The findings demonstrate a moderate to strong positive correlation between these variables, indicating that posts with higher engagement are often associated with higher toxicity scores. This trend was observed consistently across the analyzed data, supporting the hypothesis that toxic or controversial content tends to garner more attention.

To answer the research question directly: **Yes, there is a correlation between high engagement and toxicity**. The data indicates that posts with higher toxicity levels are more likely to attract increased user interaction in the form of comments and upvotes. However, this correlation should not be interpreted as causation, as multiple factors, including platform dynamics and algorithmic amplification, likely contribute to this relationship.

Looking ahead, there are several promising directions for extending this work. Developing custom sentiment analysis models tailored to the linguistic nuances of platforms like 4chan and Reddit would enhance accuracy. Expanding the analysis to include real-time tracking, multilingual data, and cross-platform comparisons could provide deeper insights into online discourse trends.

Ultimately, this project underscores the importance of understanding toxic behavior and sentiment dynamics on online platforms. Such insights have the potential to inform platform moderation strategies, combat the spread of harmful content, and contribute to broader discussions about digital ethics and community management.

8 GITHUB REPOSITORY

- **Project Implementation:** https://github.com/2024-Fall-CS-415-515/project-3-implementation-sgm
- **Final commit hash of the above repo:** 3b1110a9620492374a4d41542b9c696e39c0d8d6

REFERENCES

[1] **4chan API documentation:** *https://github.com/4chan/4chan-API.*
[2] **Official Reddit API documentation:** *https://www.reddit.com/dev/api/.*
[3] **Faktory Official Github Page:** *https://github.com/contribsys/faktory.*
[4] Moderate Hatespeech API: *https://moderatehatespeech.com/.*
[5] Flask: *https://flask.palletsprojects.com/.*
[6] Plotly: *https://plotly.com/.*
[7] Chartjs: *https://www.chartjs.org/docs/latest/.*