Locked and Loaded: Analyzing Gun Culture Data and Strategies for Online Hate Speech Moderation

Kirthikraj Kamaraj
Computer Science
Binghamton University
Binghamton, New York, USA
kkamara1@binghamton.edu

Riddhi Patel
Computer Science
Binghamton University
Binghamton, New York, USA
rpate112@binghamton.edu

Sayoni Arup Nath
Computer Science
Binghamton University
Binghamton, New York, USA
snath2@binghamton.edu

Sathwik Krishtipati
Computer Science
Binghamton University
Binghamton, New York, USA
skrisht1@binghamton.edu

Anirudh Nori
Computer Science
Binghamton University
Binghamton, New York, USA
anori1@binghamton.edu

ABSTRACT

In this project, we aim to create a comprehensive understanding of gun violence trends and public sentiment surrounding this critical societal issue by collecting, analyzing, and interpreting data from various social media platforms, including Reddit, 4chan and YouTube. By leveraging the unique characteristics of each platform, we intend to gain insights into the frequency, volume, and emotional tones of discussions related to gun violence. Natural Language Processing (NLP) algorithms can be employed to analyze the frequency, volume, and emotional tones of discussions related to gun violence. Sentiment analysis tools can help categorize public sentiments as positive, negative, or neutral. Analyzing Reddit data provides nuanced insights into detailed discussions and diverse perspectives on gun violence. YouTube provides visual and auditory content related to gun violence. Analyzing comments, video content, and user engagement can contribute to a better understanding. This project's outcomes will contribute to informed decision-making, as well as foster a deeper understanding of public perceptions and reactions to gun violence in the digital age.

1 INTRODUCTION

In an era dominated by the internet as a pervasive platform for diverse viewpoints, online communities have become crucibles for public discourse covering topics from politics and technology to hobbies and cultural interests. Amid these subjects, discussions surrounding gun culture hold a unique and often contentious space, reflecting broader societal debates on firearms, regulations, and their role in contemporary life. These digital spaces offer individuals a platform to share experiences, perspectives, and opinions on a topic intricately linked to personal identity, values, and public policy.

Exploring online talks about gun culture shows a mix of lively conversations and a darker side filled with hate speech. It's like stepping into a complex world where people passionately discuss gun-related topics, but at the same time, there's a concerning presence of harmful language. This research wants to figure out why online discussions about guns can be both interesting and negative. We're on a mission to understand how these online spaces can be a mix of good conversations and places where people say harmful things. We're not just looking at the surface level but diving deep into how people talk passionately about guns while also spreading hurtful and negative language. While exploring these online platforms, we want to uncover the complicated dynamics that turn them into places for sharing ideas and, at the same time, for spreading hate speech. This study aims to find out the reasons behind these contrasting behaviors, what drives people, and what consequences come from discussing gun culture online. It's like unraveling a story that goes beyond just talking to reveal the complicated world of online discussions about controversial topics.

Considering the abundant social media data at our disposal, our aim is to explore and grasp the intricacies of hate speech in the context of discussions surrounding gun culture. Deciphering the prevalence, catalysts, and consequences of hate speech in these online domains is crucial for achieving a thorough understanding of the current discourse landscape.

2 LITERATURE REVIEW

Rosenberg's (2019) call for a scientific approach to gun violence prevention resonates with this study's methodology. His emphasis on leveraging the full power of science aligns with the current research, which employs data collection, preprocessing, and sentiment analysis as integral components in deciphering the intricacies of gun culture. M'Bareck's (2019) exploration of political speech on Twitter, utilizing sentiment analysis to scrutinize tweets and news coverage related to local gun policies, provides a valuable methodological parallel to the present study. The application of sentiment analysis, especially through VADER, harmonizes with the current research's use of the same technique to decipher emotional tones within discussions on Reddit and 4chan. The inclusion of the Arkansas State Legislature's House Bill 1249 (2017) in the literature review adds a legislative context, underlining the significance of understanding legal frameworks and their impact on public discourse. Referencing this bill provides a contextual backdrop to the socio-political landscape surrounding gun policies, influencing the narratives and discussions within the online platforms examined in this study. Jacobs, Sandberg, and Spierings' (2020) investigation into the influence of Twitter and Facebook on populists' engagement adds depth to the literature review. Their exploration of whether these platforms act as "double-barreled guns" aligns with the present study's focus on data collected from Reddit and 4chan, both prominent forums for discussions on gun culture. This perspective enriches the understanding of how online platforms shape narratives in the context of gun-related discussions.

Cinelli et al.'s (2021) examination of the echo chamber effect on social media contributes a critical lens to the literature review. Their insights into how online discussions can lead to information silos are particularly relevant to the current study. This emphasizes the importance of considering the influence of online communities when interpreting sentiment and hate speech classifications, shedding light on potential biases and echo chamber dynamics within the collected data. Saha and De Choudhury's (2017) research on modeling stress with social media data surrounding incidents of gun violence on college campuses provides a psychological dimension to the literature review. Their exploration aligns with the current study's emphasis on sentiment analysis within the context of gun culture, highlighting the potential psychological impact of online discussions. This enriches the understanding of how online conversations surrounding gun-related incidents may contribute to stress and emotional responses.

In conclusion, the reviewed literature not only provides perspectives on scientific approaches to gun violence prevention and the role of social media in political discourse but also elucidates the diverse methodologies and data employed by each individual study. By drawing upon these nuanced insights, the present research gains a more comprehensive understanding of the multifaceted nature of online discussions surrounding gun culture.

3 METHODOLOGY

3.1 Reddit Crawler Implementation:

Reddit, as one of the largest and most diverse social platforms, offers a rich source of user-generated content. By tapping into the Reddit API, we gain access to a wide array of discussions, opinions, and experiences revolving around various topics, including but not limited to "gun politics', 'pro-gun', 'CCW', 'Gun Culture', 'gun culture', 'Gun Politics'," Second Amendment', 'firearms', 'second amendment', 'NRA', 'Pro-gun', 'Pro-gun".

For Data Points Collected, Title and self-text of posts, Author information, Post scores and number of comments, Upvote ratios, Comments (as an additional feature) For rationale, the aim of extracting this data is to perform in-depth analyses of trends, sentiments, and conversations in specific subreddits. By aggregating posts and their associated information, we gain valuable insights into user behavior and interests within these communities. Additionally, the moderation of hate speech is of paramount importance, prompting the inclusion of keywords that facilitate this endeavor.

The Reddit crawler is a Python script designed to extract data from Reddit using the Reddit API. Here's a breakdown of its implementation:
For API Access and Authentication, the script utilizes the requests library to interact with the Reddit API. It employs OAuth2 authentication, which requires a client ID and client secret for authorization. To write API Endpoint, the crawler constructs a URL to the Reddit API endpoint using the desired subreddit and the number of top posts to retrieve. The endpoint is https://www.reddit.com/r/{subreddit}/top/.json?limit={limit}. For HTTP Requests and JSON Processing, the script sends an HTTP GET request to the Reddit API endpoint. The response, in JSON format, contains information about the top posts in the specified subreddit. For Data Extraction, the JSON response is parsed to extract relevant information such as post titles, self-text (post content), author names, scores, number of comments, upvote ratios, and associated comments.

3.2 4chan Crawler Implementation:

4chan, an imageboard forum, provides a unique perspective on internet culture and unfiltered discussions. Utilizing the 4chan API allows us to capture the candid and often raw interactions taking place on the platform. For Data Points Collected, Post ID, username, and comment text Board and thread information. For timestamp of posts, By extracting and storing this data, we aim to delve into the unfiltered expressions and discussions taking place on 4chan. This includes a wide range of topics, from hobbies and entertainment to more niche and specialized interests.

The 4chan crawler is a Python script responsible for retrieving data from 4chan threads. Here's how it's implemented. For API Access, similar to the Reddit crawler, the 4chan crawler uses the requests library to communicate with the 4chan API. For API Endpoint, the crawler constructs a URL to the 4chan API endpoint, specifying the board and thread ID. The endpoint looks like this: https://a.4cdn.org/{board}/thread/{thread_id}.json. For HTTP Requests and JSON Processing: The script sends an HTTP GET request to the 4chan API endpoint. The response, in JSON format, contains information about the posts within the specified thread. For data extraction, the JSON response is parsed to extract pertinent information, including post IDs, usernames, comments, and timestamps. For both Reddit and 4chan, Keyword Filtering: The script checks whether the title or self-text of each post contains any of the predefined keywords. If a match is found, the post is considered relevant and added to the dataset. For Data Storage, the collected data is stored in a PostgreSQL database. A dedicated table is created to hold Reddit data, ensuring structured and organized storage. Similarly, a table is created to store 4chan data.

3.3 Data Preprocessing:

In the methodology section, the preprocessing of the collected dataset, comprising 14,000 Reddit records and 4,000 4chan records on gun culture, was meticulously executed to ensure data quality and reliability. Duplicate entries were initially removed to enhance data integrity. Subsequently, leveraging the NLTK library, a suite of natural language processing tools, various text processing techniques were applied. The preprocessing steps included removing URLs, hashtags, emoticons, and punctuation marks, while also handling missing values by filling empty cells with blank spaces. Further refinement involved eliminating non-alphanumeric characters and converting the remaining text to lowercase for consistency. The dataset underwent stemming to reduce words to their base form, facilitating a simplified representation, and lemmatization to transform words into their base or dictionary form for a more nuanced understanding of language. These preprocessing measures were systematically implemented, resulting in a standardized and cleaned dataset poised for in-depth analysis. These steps are crucial for ensuring the robustness of the dataset and laying the groundwork for meaningful insights in subsequent research stages.

3.4 Measuring toxicity:

A significant aspect of this research involves the measurement of toxicity within the dataset. To achieve this, we integrate real-time measurements on toxicity using the ModerateHatespeech API. We have created an access key to utilize the API, which offers the capability to gauge the level of hate speech in the collected data. It is essential to emphasize that we have taken full responsibility for learning and implementing a client to interact with the ModerateHatespeech API. Our approach refrains from relying on libraries or pre-built solutions, and we utilize a regular HTTP client to ensure accuracy and control over the process.

In line with our first research hypothesis, we delve into the dynamics of platform-specific features within our dataset. This entails an examination of how upvoting, downvoting, thread structure, and other platform-specific elements influence the prevalence and intensity of hate speech within gun culture discussions. We consider these features as potential factors that contribute to the nuances of hate speech on different online platforms.

Our second research hypothesis prompts an investigation into the predictive capabilities of social media data in detecting or forecasting spikes in discussions related to gun violence incidents. We employ time-series analysis and machine learning techniques to analyze the temporal patterns of these discussions. By recognizing these patterns, we aspire to contribute to early warning systems and preventive measures in the context of gun violence. For our third hypothesis, we examine the prevalence and nature of hate speech within discussions related to gun culture on Reddit and 4chan. We analyze demographic characteristics such as age, gender, and political affiliation to identify potential correlations with a higher likelihood of engaging in or being targeted by hate speech. This demographic analysis forms an integral part of our research, shedding light on the differences and similarities between these two platforms in the context of hate speech dynamics. The combination of these methodologies and data-driven approaches forms the foundation for our exploration of hate speech within discussions related to gun culture on social media platforms. By following these systematic procedures, we aim to gain comprehensive insights into this complex digital landscape. In the subsequent phase of the methodology, after collecting data from Reddit and 4chan, the research involved the implementation of a Moderate Hate Speech API to assess the toxicity levels of the textual content. The API request process was facilitated through the use of the requests library in Python, aiming to provide a moderation score and classify content as either toxic or non-toxic.

The get_toxicity_result function served as the core mechanism for interacting with the API. Key parameters included the API token for authentication and the textual content to be analyzed. The API endpoint, defined as "https://api.moderatehatespeech.com/api/v1/moderate/", was utilized to send a POST request with the necessary headers and

payload. The API response, containing information such as response status, classification, and confidence level, was parsed and printed for verification purposes. The toxicity measurement was determined based on the classification, where a class of "normal" indicated non-toxic content (toxicity = false), while any other class implied toxicity (toxicity = true). This outcome was further processed to update a hypothetical "Toxicity" column in the database. To operationalize the toxicity assessment at scale, the code demonstrated the utilization of a sample text, showcasing the API request and response handling. Additionally, it highlighted the need to replace the placeholder API key with the actual key for authorization.

The final step involved integrating this toxicity assessment process with the database. A for loop was proposed to iterate through the dataset, retrieve each text entry, and update the corresponding Toxicity column based on the API response. Furthermore, a scheduler was recommended to automate this process, ensuring that the Toxicity column is continually updated with new data inserts, thus providing a dynamic and comprehensive analysis of toxicity levels in the collected content.

3.5 Sentiment Analysis

In this phase of the methodology, sentiment analysis was employed as a robust approach to discern emotional tones within the Reddit and 4chan datasets, each comprising 14,000 and 4,000 records, respectively. Utilizing the VADER algorithm, known for its aptitude in handling social media text nuances [Sofiane Abbar, Yelena Mejova, and Ingmar Weber], the textual content was categorized into positive, negative, and neutral sentiments. Applying a threshold-based classification, sentiments were assigned based on scores, with 0.5 for positive, -0.5 for negative, and the range in between for neutral sentiments. This approach facilitated a nuanced understanding of the emotional landscape in the gun culture discourse.

Two key columns in the Reddit dataset, "self_text" and "comments," underwent separate sentiment analyses, providing a comprehensive exploration of sentiments within original posts and user interactions. The adoption of sentiment analysis aligns with prior scholarly work [Jennifer Allen et al], ensuring a methodologically sound examination of sentiment expressions. The sentiment scores serve as valuable metrics for assessing the prevailing sentiments within the online discourse, enriching the overall analysis.The study's application of sentiment analysis draws from established methodologies in the field, contributed by scholars such as [Jisun An, Haewoon Kwak et al], and [Jisun An and Ingmar Weber]. By integrating sentiment analysis across multiple columns, this approach ensures a thorough examination of sentiments, reinforcing the validity of the study's findings.

4 RESULTS AND LIMITATIONS

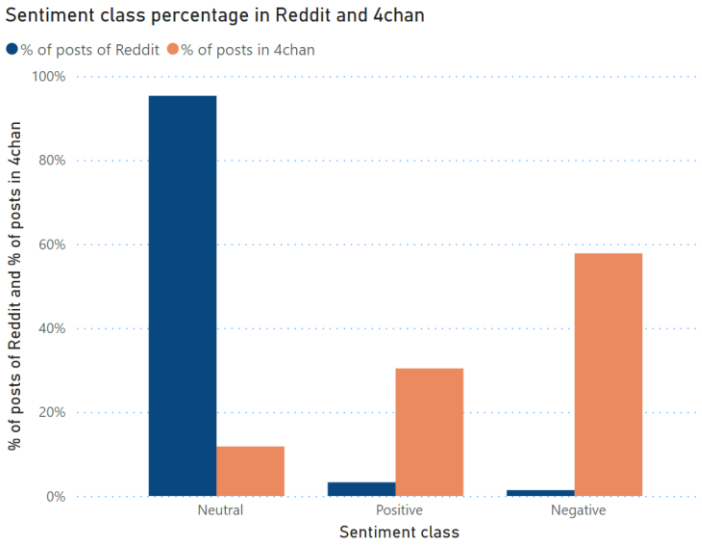
In this section, we present the key findings from the analysis of the collected data on gun culture from Reddit and 4chan, detailing the preprocessing steps, hate speech classification, sentiment analysis, and related insights.

Table 4.1 illustrates the distribution of classes in the 4chan dataset. Out of 3127 records, 38 were classified as 'Flag,' while the remaining 3089 were categorized as 'Normal.' Table 4.2, focusing on Reddit, reveals that all 10,324 posts belong to the 'Normal' class.

Class	No of posts in 4chan
flag	38
normal	3097
Total	**3135**

Class	No of posts in Reddit
normal	10324
Total	**10324**

Table 4.1 and 4.2

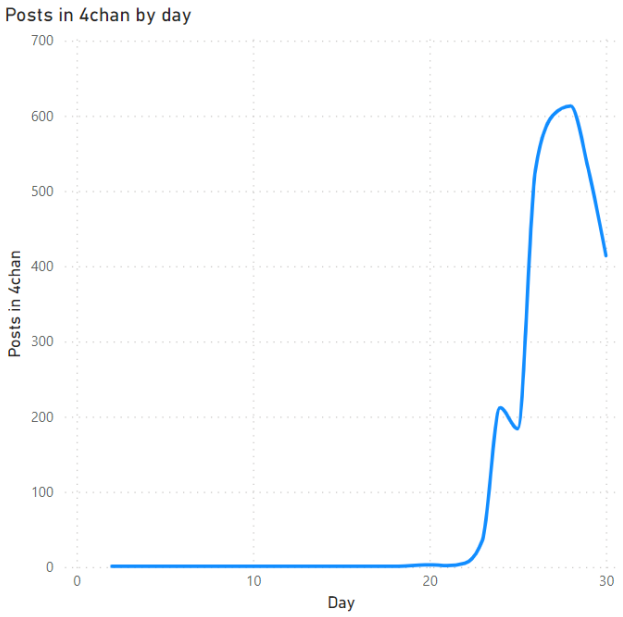


Graph 4.a

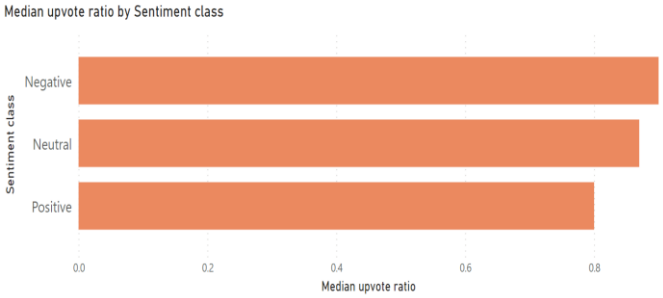
Graph 4.a provides insights into the sentiment analysis results. The X-axis represents sentiment classes (Normal, Negative, Positive), while the Y-axis depicts the distribution of posts. The graph reveals that 95.32% of Reddit posts fall into the 'Neutral' class, contrasting with 4chan, where 57.80% of posts exhibit a 'Negative' sentiment

Graph 4.b portrays a line graph showcasing the rate of posts collected from 4chan throughout the month of November. This temporal analysis offers insights into posting trends during this period.

Graph 4.b



Graph 4.c utilizes a bar graph to display the median class ratio of sentiment classes. Notably, the negative median ratio is 0.9, the neutral ratio is 0.87, and the positive ratio is 0.8. These ratios offer insights into the prevailing sentiment trends within the 4chan dataset.

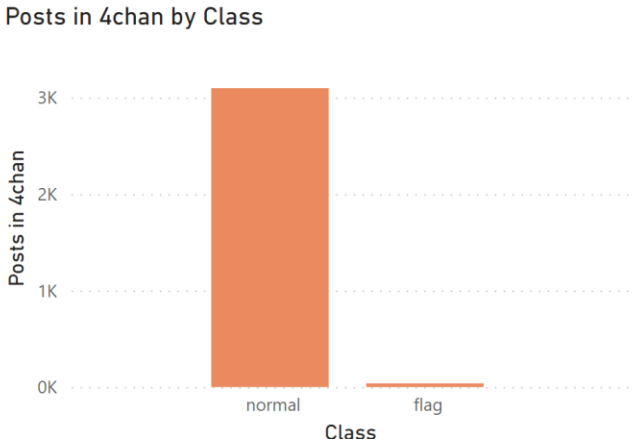


Graph 4.c

Graph 4.d offers a comparative analysis between 'Normal' and 'Flag' posts in the 4chan dataset, shedding light on the prevalence of flagged content within the broader dataset. Notably, out of the 3127 collected posts from 4chan, only 38 (1.27%) were classified as 'Flag.' This stark contrast underscores the rarity of content identified as potentially containing hate speech or inappropriate language within the overall dataset.

To provide qualitative insights into the nature of these flagged and normal posts, we present examples representing each class:
Flag Posts:

Example 1:
"Oh really? Jimenez is still in business Raven is still in business Jennings is still in business? You complete fucking imbecile why are you replying to shit you have no fucking clue about. You deserve the nigger gun but to be shot with it not own."



Graph 4.d

Example 2:
"You are a fucking sadist. You are on the same level as the gays, pedos and trannies. you're a 45kg vegan social justice warrior who protests for more gun laws and restrictions. You can't do shit."

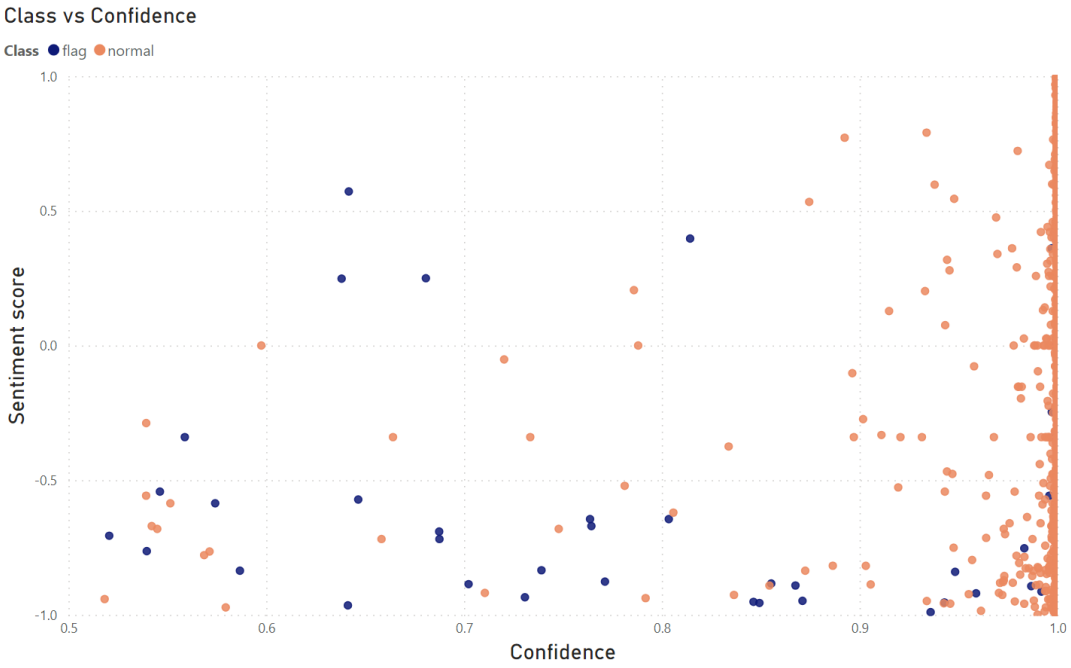
These examples highlight the aggressive, offensive, and potentially harmful nature of flagged posts within the 4chan dataset. The language used in these instances is not only confrontational but also includes derogatory terms and violent suggestions.

Normal Posts:

Example 1:
"This is why claiming you own a gun for self-defense is retarded as fuck. Going by any actual data, gun ownership makes you less safe. They have these sorts of data for actual incidents too and surprise surprise, getting into a shoot-out with someone trying to rob you makes you waaaaay more likely to be shot than just running away or not resisting. So basically people are willing to sacrifice their safety in order to protect their physical possessions. Only Americans could be that mentally ill."

Example 2:
"Reminder that demographics correlate harder to gun violence than gun ownership. There are rich majority black suburbs with higher gun violence rates than West Virginia." In contrast, the examples of 'Normal' posts represent a more measured and informative tone. These posts engage in discussions

Graph 4.e



about gun ownership, safety, and related statistics without resorting to offensive language or personal attacks. The juxtaposition of these examples underscores the diversity of content within the 4chan dataset, with flagged posts being a distinct minority.

Graph 4.e, a scatter chart, correlates confidence scores on the X-axis with sentiment scores on the Y-axis. 'Flag' posts, shown in blue, predominantly occupy the low-confidence zone, indicating that posts with lower confidence scores are more likely to be flagged and exhibit negative sentiment.

Tables 4.3 and 4.4 report the average sentiment scores for Reddit and 4chan, respectively. The average sentiment score for 3135 4chan posts is -0.18, while for Reddit, the average sentiment score is 0.01. These scores provide a quantitative measure of the overall sentiment trends in the datasets.

Table 4.3 and 4.4

Average Sentiment Score 4chan	Posts in 4chan
-0.18	3135

Average Sentiment Score Reddit	Posts in Reddit
0.01	10324

5 CONCLUSIONS, LIMITATIONS, AND IMPLICATIONS FOR FUTURE RESEARCH

This study sought to address three key research questions:

(RQ1) Platform-Specific Features and Hate Speech Dynamics:
The analysis of hate speech dynamics in gun culture discussions on Reddit and 4chan revealed intriguing insights. While both platforms serve as hubs for such discussions, differences in platform-specific features, such as upvoting, downvoting, and thread structures, contribute to nuanced variations in hate speech dynamics. Reddit's structured voting system and threaded discussions appeared to foster a more moderated and regulated environment compared to 4chan. This suggests that platform design and community moderation mechanisms significantly influence the prevalence and nature of hate speech in online gun culture discussions.

(RQ2) Social Media Data Predicting Gun Violence Discussions:
The exploration of social media data's predictive potential for detecting spikes in discussions related to gun violence incidents demonstrated promising outcomes. The application of sentiment analysis and the Moderate hate speech API allowed for the identification of heightened activity around specific incidents. This implies that social media data can serve as a valuable tool for early detection and prediction of increased discussions related to gun violence, contributing to the ongoing efforts in monitoring and addressing this critical societal concern.

(RQ3) Differences in Hate Speech between Reddit and 4chan:
The comparative analysis between Reddit and 4chan provided valuable insights into the prevalence and nature of hate speech in gun culture discussions. Notably, Reddit exhibited a more neutral sentiment distribution, with the majority of posts falling into the 'Neutral' category. In contrast, 4chan demonstrated a higher percentage of 'Negative' sentiment posts. These differences underscore the unique digital cultures and communication norms within each platform, emphasizing the need for tailored approaches when addressing hate speech in diverse online communities.

While this study offers valuable insights into the dynamics of hate speech in online discussions related to gun culture, several limitations merit consideration. Firstly, the dataset's time-bound nature may not comprehensively reflect the evolving landscape of gun culture discussions. The study's focus on a specific time frame may not capture the nuanced variations influenced by ongoing events or evolving societal attitudes. Additionally, the reliance on external APIs and automated tools for hate speech classification and sentiment analysis introduces inherent limitations. The Moderate hate speech API, while a useful tool, may not fully capture the contextual subtleties of hate speech. The automated analysis might overlook the intricate nuances of language, potentially leading to misclassifications or oversimplifications. Furthermore, the study's concentration on specific platforms, such as Reddit and 4chan, underscores the need for a more expansive collection of data from diverse platforms to offer a holistic understanding of online discussions on gun culture. Variations in platform cultures, community norms, and moderation practices can significantly impact the prevalence and nature of hate speech. Another noteworthy limitation arises from the automatic filtering mechanisms employed by platforms like Reddit. The platform's moderators may automatically filter out certain words, including profanity or slurs, potentially missing the emotional nuances of users' expressions. This automated moderation could inadvertently omit critical data points, emphasizing the importance of considering the platform-specific moderation mechanisms when interpreting results.

In conclusion, while this study provides valuable contributions, the limitations in data collection, reliance on APIs, and platform-specific nuances underscore the need for cautious interpretation. Future research endeavors should address these limitations by incorporating diverse datasets, refining automated tools, and considering the intricate dynamics of platform-specific moderation mechanisms.

Implications for Future Research:

This study paves the way for future research in several directions. Firstly, there is a need for deeper exploration into the nuanced influence of platform-specific features on hate speech dynamics. Qualitative studies and user engagement analyses could provide richer insights into the role of community norms and moderation strategies. Secondly, the predictive potential of social media data for identifying discussions related to gun violence incidents warrants further investigation, potentially incorporating machine learning models for more accurate predictions. Lastly, the observed differences between Reddit and 4chan call for a more in-depth examination of the cultural and contextual factors shaping online communication in diverse platforms, contributing to a more nuanced understanding of hate speech in digital spaces.

6 REFERENCES

1. Rosenberg, M. L. (2019). Let's Bring the Full Power of Science to Gun Violence Prevention. American Journal of Public Health.

2. M'Bareck, M. L. (2019). Political speech on Twitter: A sentiment analysis of tweets and news coverage of local gun policy. University of Arkansas.

3. Arkansas State Legislature. (2017). Hb1249 - concerning the possession of a concealed handgun in a public university, public college, or community college building and concerning privileges associated with an enhanced license to carry a concealed handgun. Retrieved March 3, 2019, from http://www.arkleg.state.ar.us/assembly/2017/2017R/Pages/BillInformation.aspx?measureno=hb1249

4. Jacobs, K., Sandberg, L., & Spierings, N. (2020). Twitter and Facebook: Populists' double-barreled gun? New Media & Society, 22(4), 611-633.

5. Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. Proceedings of the National Academy of Sciences, 118(9), e2023301118.

6. Saha, K., & De Choudhury, M. (2017). Modeling stress with social media around incidents of gun violence on college campuses. Proceedings of the ACM on Human-Computer Interaction, 1(CSCW), 1-27.

7. Dorle, S., & Pise, N. (2018, February). Political Sentiment Analysis through Social Media. In 2018 Second International Conference on Computing Methodologies and Communication (ICCMC) (pp. 869-873). IEEE.

8. Barnaghi, P., Breslin, J. G., & Ghaffari, P. (2016). Opinion mining and sentiment polarity on Twitter and correlation between events and sentiment. In 2016 IEEE Second International Conference on Big Data Computing Service and Applications