

# Analyzing Social Trends on Gun Violence

Kirthikraj Kamaraj  
Computer Science  
SUNY Binghamton University  
Binghamton, New York, USA  
[kkamara1@binghamton.edu](mailto:kkamara1@binghamton.edu)

Riddhi Patel  
Computer Science  
SUNY Binghamton University  
Binghamton, New York, USA  
[rpate112@binghamton.edu](mailto:rpate112@binghamton.edu)

Sathwik Krishtipati  
Computer Science  
SUNY Binghamton University  
Binghamton, New York, USA  
[skrisht1@binghamton.edu](mailto:skrisht1@binghamton.edu)

Sayoni Arup Nath  
Computer Science  
SUNY Binghamton University  
Binghamton, New York, USA  
[snath2@binghamton.edu](mailto:snath2@binghamton.edu)

Anirudh Nori  
Computer Science  
SUNY Binghamton University  
Binghamton, New York, USA  
[anori1@binghamton.edu](mailto: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, Twitter, and 4chan. 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. 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

Gun violence represents a significant and persistent challenge in society. Public discourse on this issue has extended to social media platforms, where conversations, opinions, and reactions are actively shared. To address this, our project focuses on the collection, analysis, and interpretation of data from Reddit, Twitter, and 4chan to shed light on the evolving landscape of gun violence discussions. By doing so, we aim to provide valuable insights that can inform future research, policy decisions, and public awareness campaigns.

## 2. THIRD DATASET: 4chan:

4chan is a unique and often controversial imageboard forum that serves as our third data source for understanding discussions related to gun violence. This platform stands apart from conventional social media platforms like Twitter and Reddit due to its anonymous and decentralized nature. 4chan is structured into various boards, each dedicated to specific topics or interests, where users engage in discussions through posts and comments, often with limited moderation.

For our project, we will target specific boards or threads within 4chan that contain discussions about gun violence. These discussions may include text-based comments, images, and links, all of which contribute to the platform's distinct culture and communication style. Given 4chan's

characteristics, data collection from this source presents unique challenges and opportunities. We will use the 4chan API, when available, to access these discussions and extract relevant data.

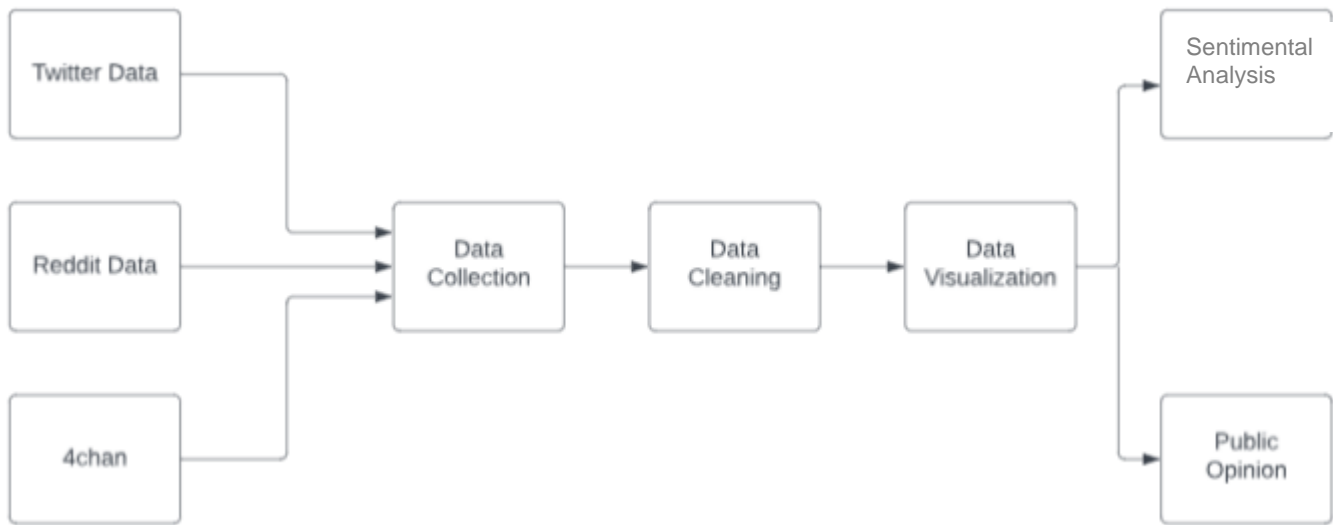
## 3. DATA SOURCES

The methodology of this project encompasses several key components. We will meticulously select our data sources, targeting specific subreddits on Reddit, relevant hashtags and keywords on Twitter, and defining criteria for data collection from 4chan. Our data collection strategy will rely on Python programming Twitter API, Reddit API and 4chan Api with web scraping techniques used where necessary. To ensure continuity, we will implement data collection scripts that run continuously. Data preprocessing will involve cleaning and normalizing text data for sentiment analysis, and collected data will be securely stored in an organized format for easy retrieval.

## 4. ANALYSIS APPROACH

Analysis Approach: Our analysis approach comprises two fundamental components. First, we will measure and analyze trends in the frequency and volume of gun violence-related discussions across the selected platforms. Secondly, we will apply sentiment analysis techniques to assess public sentiment and emotional tones present in the posts. Through these analyses, we aim to address research questions related to prevailing sentiments, platform-specific differences, and key influencing factors shaping public

perceptions of gun violence.



**Figure 1:** Data Flow Diagram for Social Media Analysis

## 5. NAPKIN MATH

In terms of the data format, we will be working with text data, which will be stored in CSV or XLSX files. Each week, we anticipate collecting around 200,000 records of text data. This means that over the course of two weeks, our dataset will grow to approximately 400,000 records. After this initial two-week period of data collection, our focus will shift towards essential preprocessing and other crucial steps in our data analysis pipeline.

Considering the volume of data we plan to collect, it's important to note that we will need a minimum of 1 gigabyte (GB) of hard disk space to accommodate this dataset comfortably. This space will allow us to store, manage, and analyze the text data effectively as we progress in our project.

## 6. REFERENCES

[1] TwitterAPI Documentation.  
<https://developer.twitter.com/en/docs/twitter-api>

[2] RedditAPI Documentation  
<https://www.reddit.com/dev/api/>

[3]4chan.org:  
<https://www.4chan.org/>