Analyzing Political Trends on Social Media

Paritosh Marathe
Computer Science
SUNY Binghamton University
Binghamton, New York, USA
pmarathe1@binghamton.edu

Piyush Rathod
Computer Science
SUNY Binghamton University
Binghamton, New York, USA
prathod@binghamton.edu

Dhanashree Borkar
Computer Science
SUNY Binghamton University
Binghamton, New York, USA
dborkar1@binghamton.edu

Parinita Chinchalikar
Computer Science
SUNY Binghamton University
Binghamton, New York, USA
pchinchalika@binghamton.edu

Ankit Patil
Computer Science
SUNY Binghamton University
Binghamton, New York, USA
apatil@binghamton.edu

ABSTRACT

In the digital era, social media is the predominant communication platform, engaging a diverse audience and facilitating public expression across various subjects. This study underscores the critical understanding of current social trends, public sentiments, and political discourse for researchers and policymakers. The primary focus is an in-depth examination of social media trends, utilizing data analytics and sentiment analysis to highlight prevailing political currents influencing the general populace. This involves analyzing extensive social media data, providing insights into the interplay between politics and public sentiment in the digital age.

Social media functions as a valuable resource, offering insights into collective sentiments on diverse topics and public perceptions of government policies, whether positive or negative. The research project aims to equip researchers and policymakers with the knowledge and tools necessary to navigate the evolving landscape of public opinion and political discourse, striving for a clearer understanding of the factors shaping our political landscape in the digital age. Through a systematic exploration of social media dynamics, this effort employs data-driven techniques to identify patterns, trends, and sentiments. The alignment of social media trends with real-world events highlights its substantial impact on public opinion. This project delves into political trends on social media, using data analytics to uncover underlying dynamics. The insights aim to enhance understanding of the evolving political landscape, providing a valuable resource for researchers and policymakers in navigating the intricacies of contemporary politics in the digital age.

KEYWORDS

Reddit, *r/politics*, Hate speech, Data collection, political posts, elections, political issues, PostgreSQL, upvote, downvote.

1. INTRODUCTION

Political news holds a central role in a nation's governance, forming a cornerstone of societal progress. The realm of political matters extends well beyond traditional venues like polling stations and televised debates. In today's digital landscape, where the influence of social media and digital platforms is all-encompassing, there exists a wealth of untapped information. Enter the age of Big Social Data, a valuable resource for researchers seeking to understand public engagement, sentiment, and behaviors. By harnessing the abundant data generated in the digital realm, we can simulate and analyze the intricate dynamics that shape public opinion and information dissemination. Our journey is guided by several compelling motivations:

1. Exploring Engaging Topics: Our primary motivation is to uncover the topics that drive online discussions. The digital sphere offers a unique opportunity to investigate what sparks conversations, whether it's contentious political issues or culturally relevant subjects. This exploration aids us in comprehending the factors that drive active online participation, revealing the issues that resonate within the digital community.

2. Issue Analysis and Framing for User Engagement: Another driving force is the exploration of how political topics are presented in online discussions to engage users. We aim to decode the strategies employed to shape public perception and discourse within the digital sphere. By understanding how political issues are framed and reframed, we gain insights into the techniques that resonate with digital audiences by gathering word cloud based on *r/politics*.

3.Study of Hate Speech and Abuse: A critical motivation is the examination of hate speech and abusive content in online posts. While the digital realm promotes free expression, it also harbors darker aspects of human behavior. Our aim is to shed light on the prevalence, origins, and impact of hate speech and online abuse. We seek to understand the consequences and implications of such behavior in the online political landscape and develop strategies to mitigate its harmful effects. Notably, platforms like Reddit and 4Chan play pivotal roles in this understanding. They serve as digital spaces where public opinions are distilled and reactions to the ever-evolving political landscape are captured. These platforms offer a unique perspective on the global pulse, providing a lens through which we can examine current affairs trends worldwide [1][5]. In our project, we undertake a comprehensive analysis of posts on Reddit and 4Chan, platforms known for their political engagement. Our objective is to provide an accurate representation of key political events from diverse perspectives. Our endeavor goes beyond observing user behaviors; it delves into the landscape of discussion topics, political affiliations, and emotional dynamics that underlie these vibrant digital communities [1][5]. Through this, we aim to gain a deeper understanding of the multifaceted dimensions of contemporary political engagement, revealing the interconnected nature of our digitally connected world.

2. DATA SOURCES AND BACKGROUND

Social Media such as SNS, weblogs, microblogging, and wikis play an increasingly important role in shaping political communication in the US and around the world.[2] An information network made up of short messages (including multimedia posts) from all over the world wherein Reddit is a social news website and forum where content is socially curated and promoted by site members through voting. It contains a unique upvote and downvote system which helps the users to manage the public data [3]. Table1 shows the subreddits sum of up votes and down votes

Subreddits	Sum of score	Sum of ups	Sum of downs
AmericanPolitics	108083	108083	0
conservatives	325746	325746	0
moderatepolitics	3095246	3095246	0
political	33224	33224	0
politics	11940809	11940809	0
Grand Total	15503108	15503108	0

Table1

Along with Reddit, 4Chan is also another social forum used for public opinion and interaction among a huge number of people based on their thoughts [4]. To our knowledge, empirical studies addressing the extent of hate speech in alt-right fringe communities have exclusively relied on automated dictionary-based approaches, estimating the amount of hate speech by identifying posts that contain hateful terms. Focusing on *4chan/pol/* use the hate base dictionary to assess the prevalence of hate speech in the "Politically Incorrect" board. They find that 12% of posts on *4chan/pol/* contain hateful terms. However, 4chan generally seems to be more hateful than other social media platforms [8]. The following hate terms appeared most "*nigger*," "*faggot*," and "*retard*" are indicative of racist, homophobic, and ableist sentiments and suggest that people of color, the lesbian, gay, bisexual, transgender and queer or questioning (LGBTQ) community, and people with disabilities might be recurrent victims of hate speech [8].

Our data crawler is essentially gathering information from both 4Chan and Reddit platforms, Fig 1 storing this data in a PostgreSQL database, and then further processing and preparing it for display on our dashboard wherein we started collecting a small amount of data which included over *100-200 posts per day* continuously for *3 days*. We calculated an estimate of collecting over *90-100 GB* of data over a period of 2 months, the estimate may vary as the total number of subreddits, and boards may be increased over the period depending on the political trend analysis. If required, we may introduce new sub reddit's and other API's while implementation to achieve goal.
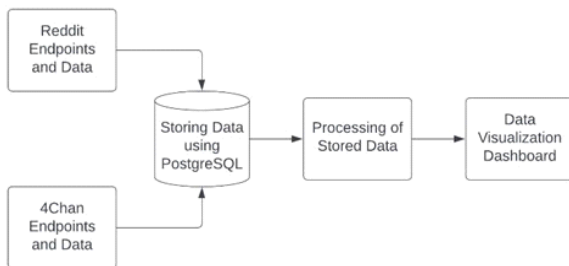

Fig 1

2.1 DATASET

During the exploration of engaging topics and issue analysis, we obtained a word cloud based on the *r/politics* subreddit. This data proved instrumental in discerning how political issues are presented and capture the attention of a larger audience. Furthermore, it facilitated our examination of hate speech and abusive language. The given below figure Fig 2 describes the word cloud based on *r/politics*.

Fig 2

We have designed Table 2 "reddit_tb table schema", which has columns below in it. The following table explains each attribute in table.

id	Id of the post
subreddit	Name of the subreddit
title	Title of the post
selftext	Post description
upvote_ratio	Ration of upvote:downvote
ups	Upvotes
downs	DownVotes
score	Not Using right now
created	Timestamp of a post
text	Not Using right now

Table 2

By using the above data we have plotted graph based on many parameters. Fig 3 shows bar graph of comments per subreddits. Fig 4 shows politics post binned daily. Fig 5 shows Hate speech per subreddit. Fig 6 shows comments binned hourly in bar graph. Fig 7 shows hate speech percentage per subreddit.
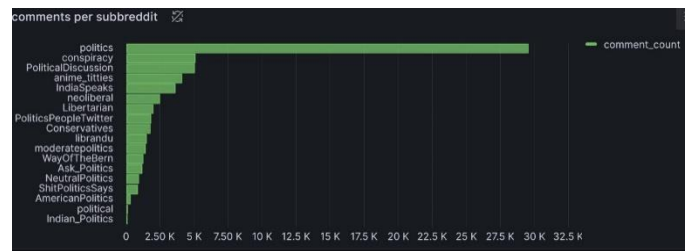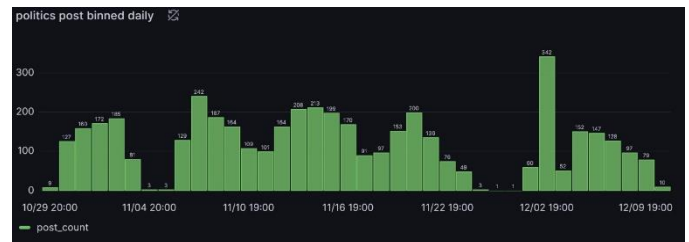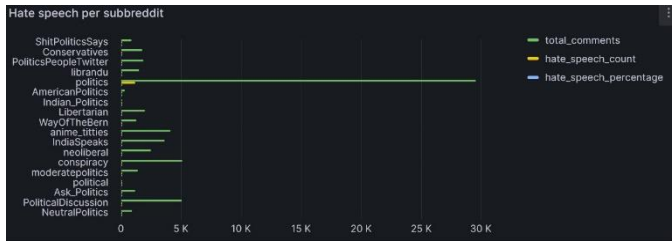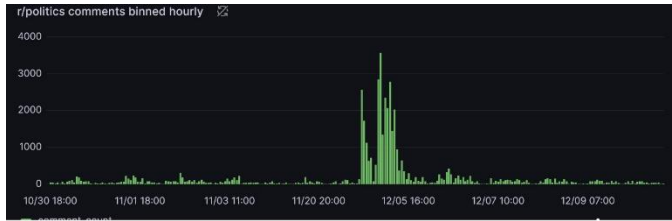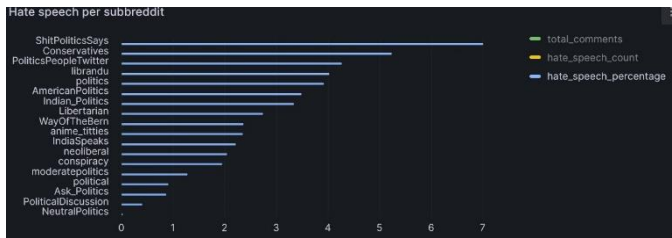

Fig 3

Fig 4

Fig 5

Fig 6

Fig 7

3. METHODOLOGY

We have divided this project phase in three steps based on the project description 2. These steps will be important to explore the research objectives in a correct manner. The steps comprise of Data collection, Processing and Result analysis.

3.1 DATA COLLECTION

We started by running a Python script on the virtual machine to collect data based on the trending political news and current situations going on along with the view of the people through the comments posted. The data has been collected from the Reddit API and 4chan API depending on what post and comments it has. It collects data in form of comments, number of upvotes, downvotes and overall score based on previous two values. We are using PostgreSQL to collect data without any extra columns being collected.

3.2 DATA VISUALIZATION

Data can be shown, or presented using the values collected such as upvotes, downvotes and ratio showing how people react and discuss. It shows what kind of political trend makes the common people engage and put their opinion on the social media.

3.3 DATA PROCESSING

The collected data (reddit or 4chan posts) contains many issues which need to be addressed. When we first retrieved the data, it contains lot of URLs which are referring to other posts. Additionally, removal of emoticons and stop words is also necessary because it will hinder the analysis results. In data, we found out that there is use of curse words like "write the curse words here" So to bifurcate the data we use these curse words as key identifiers texts.

4. PROBLEM FACED

In the development phase, we discovered a restriction on the number of comments retrievable per subreddit. Specifically, Reddit limits the storage of the latest 1000 comments per subreddit. Additionally, the Reddit API imposes a constraint of 10 queries per minute (QPM). However, this limitation can be addressed through authentication, which elevates the restriction to 100 QPMs. This enhancement is slated for

implementation in the next phase. Sometimes the virtual machine unexpectedly terminates the screen function for active processes.

5. DISSCUSSION AND CONCLUSION

We gathered comments from posts obtained through the Reddit API. Subsequently, we analyzed and derived insights into public opinion regarding the political crisis, inflation, and the impact of the upcoming election. This analysis involved filtering relevant keywords to identify trends and updates.

1. In sentiment analysis, we determined the types of posts receiving more upvotes and identified the keywords that encourage user interaction.
2. Analyzing the word cloud from *r/politics*, we examined the discussions surrounding specific political issues, their framing, and prioritization in online political posts (Issue analysis).
3. Additionally, we investigated the occurrence of hate speech, harassment, and online abuse within political posts (Online Hate Speech and Harassment).

This approach allowed us to address all three proposed research areas comprehensively.

6. REFERRENCES

[1] Analyzing voter behavior on social media during the 2020 US presidential election campaign https://doi.org/10.1007/s13278-022-00913-9
[2] Reddit Glossary Documentation. https://www.reddit.com/r/TheoryOfReddit/wiki/glossary
[3] Reddit Documentation. https://www.reddit.com/dev/api/
[4] 4Chan API Documentation. https://github.com/4chan/4chan-API
[5] Stance Detection on Social Media: State of the Art and Trends https://arxiv.org/pdf/2006.03644.pdf
[6] Mapping Publication Trend of Political Parties Campaign in Social Media: A Bibliometric Analysis
https://www.researchgate.net/publication/36240
[7]2479_Mapping_Publication_Trend_of_Political _HYPERLINK
"https://www.researchgate.net/publication/362402479_Mapping_Publication_Trend_of_Politic al_Parties_Campaign_in_Social_Media_A_Bibliometric_Analysis"Parties_Campaign_in_Socia l_Media_A_Bibliometric_Analysis
[8] Social Media and political communication: a social media and analytics framework
[9] Assessing the Extent and Types of Hate Speech in Fringe Communities: A Case Study of Alt-Right Communities on 8chan, 4chan, and Reddit https://journals.sagepub.com/doi/full/10.1177/205630512110529