

Tracking Political Sentiments: Analyzing Election-Driven Discussions on Reddit and 4chan

Asmita Singh
Binghamton University
Binghamton, NY, USA
asingh14@binghamton.edu

Chinmay Purandare
Binghamton University
Binghamton, NY, USA
cpurandare@binghamton.edu

Sanika Shinde
Binghamton University
Binghamton, NY, USA
sshinde2@binghamton.edu

Abstract

This proposal document for our project covers various requirements such as data sources, API endpoints, system architecture, third party libraries, data measurements and analysis ideas, estimated data storage etc.

Keywords

Reddit API, 4Chan API, PostgreSQL, Data Collection, Data Pipeline, Data Measurement and Analysis, Plotting, Visualization

1 Introduction

Our study will focus on the upcoming 2024 US election, which will be the 60th quadrennial presidential election. The goal of our project is to characterize netizens' reactions on the ongoing political campaigns, issues, debates etc. that occur during the days leading up to the general election. During this period, online communities often become more active because of political events such as rallies, debates or statements made by presidential candidates. These events trigger positive and negative responses toward the candidates, influencing public opinion. We plan to create a data pipeline to collect data related to these reactions from platforms like Reddit and 4chan. We will collect live data from Reddit and 4chan which we believe will give us some interesting insights, especially considering the differences that Reddit and 4chan have in makeup of their communities. The outcome of our study is an analysis of different political ideologies of Americans.

2 Data Source

The data for this study will be collected by our crawler, running on the 128.226.29.138 VM, from two online platforms Reddit and 4chan, both known for their active political communities. Specifically, we will collect posts and comments from some or all of the below subreddits:

<https://www.reddit.com/r/Liberal/>
<https://www.reddit.com/r/Conservative/>
<https://www.reddit.com/r/politics/>
<https://www.reddit.com/r/NeutralPolitics/>
<https://www.reddit.com/r/KamalaHarris/>

For 4chan, we will collect data from the board /pol (Politically Incorrect). Both Reddit and 4chan have APIs which will be used for data collection.

2.1 API endpoints

The Reddit API endpoints which we will use are:

https://www.reddit.com/dev/api#GET_hot
https://www.reddit.com/dev/api#GET_new
https://www.reddit.com/dev/api#GET_{sort}

https://www.reddit.com/dev/api#GET_r_{subreddit}_about

The 4chan API endpoints are:

https://github.com/4chan/4chan-API/blob/master/pages/Endpoints_and_domains.md#api-endpoints

3 System Architecture

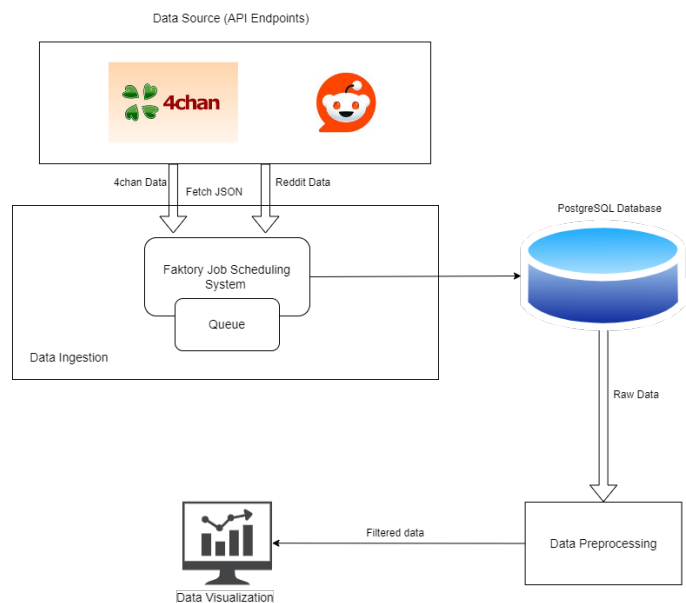


Figure 1: System Architecture

3.1 Third party libraries

- NLTK
- NRCLex
- Google Perspective API [1]
- WordCloud

4 Data measurements and analysis

We plan to use Machine Learning on the Reddit and 4chan datasets for identifying sentiments (positive or negative), emotions, frequently used terms, Named Entity Recognition, generating WordClouds etc. Perspective API will be used for comparing hate speech between Reddit and 4chan communities. Comparing user reactions in different subreddits will help us examine how Americans of different political ideologies respond to the same political events.

5 Estimated weekly data collection

We checked the subreddits in our data source list to see how many average daily posts they received to come up with a rough estimation of weekly user posts: depending on the size of the subreddit, the weekly average posts ranged from 490 (r/KamalaHarris) to 2450 (r/politics).

For 4chan, due to the ephemeral nature of threads on 4chan's /pol board, it was difficult to estimate the number of weekly threads so we used an online tool 4stats.io which gave us an estimate of 21168 threads/week and 700K replies/week.

6 Citations and Bibliographies

We referred An Early Look at the Gettr Social Network[2]

References

- [1] Perspective API. [n. d.]. ([n. d.]). <https://www.perspectiveapi.com/>
- [2] Emiliano De Cristofaro Savvas Zannettou Gianluca Stringhini Pujan Paudel, Jeremy Blackburn. 2021. AnEarly Look at the Gettr Social Network. (Aug. 2021), 1–10. <https://doi.org/10.48550/arXiv.2108.05876>