

IMPACT OF RACISM ON US ELECTIONS 2020

Shashank Bengaluru Srinivasa
sbengal2@binghamton.edu
State University of New York at
Binghamton
Binghamton, New York

Aravind Reddy Yenugula
ayenugu1@binghamton.edu
State University of New York at
Binghamton
Binghamton, New York

Darshan Doddaghatta
ddoddag1@binghamton.edu
State University of New York at
Binghamton
Binghamton, New York

ABSTRACT

Race and ethnicity have always played a role in politics around the world, in the same order they influencing the politics in a diverse and multicultural country like USA comes as no surprise. In the light of recent events triggered by the death of **George Floyd**, it surely does seem like it is going to have its effect on the upcoming US Elections 2020.

In this project we intend to gain an overview of the general public behind these events, its influence on their political stand and through all this try to predict the result of the US elections 2020. What better option than social media for this ? Therefore, we shall use social media sites such as twitter and reddit to collect the required data. We intend to perform a lexicon based sentiment analysis over the data and try to predict the outcome of the elections.

ACM Reference Format:

Shashank Bengaluru Srinivasa, Aravind Reddy Yenugula, and Darshan Doddaghatta. 2018. IMPACT OF RACISM ON US ELECTIONS 2020 . In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/1122445.1122456>

1 DATA SOURCE

Data shall be collected from two sources

Twitter

- Twitter's **sample stream API** is used to collect the data such as tweets/retweets.
- The required data would be obtained using multiple hashtags such as:
 - #blacklivesmatter
 - #racism
 - #whatmatters2020
 - #icantbreathe
 - #USElections
 - #2020elections
 - #trumpvsbiden
- The data is usually a tweet object. We shall apply filters and collect data only from the relevant fields in the tweet object.

Reddit

- Reddit **search** API is used to collect the data such as Reddits/subreddits.
- The reddit data is queried by filtering it using relevant keywords.

2 DATA COLLECTION

2.1 Twitter

Twitter has rich web APIs and large amounts of data. As collecting data using sample stream API is mandatory we shall be using the same for the twitter data collection.

The sampled stream endpoint delivers a roughly 1% random sample of publicly available Tweets in real-time.

This streaming endpoint delivers **Tweet objects** through a persistent HTTP GET connection. You can connect one client per session, and can disconnect and reconnect no more than 50 times per 15 minute window

2.2 Reddit

Reddit is equipped with well designed APIs that are quite sufficient for our project

Reddit's search API is used to obtain Reddits/subreddits using specific keywords related to our project.

Reddit APIs return **JSON objects** as response to requests. We can filter the fields in the JSON objects and get the required data (text).

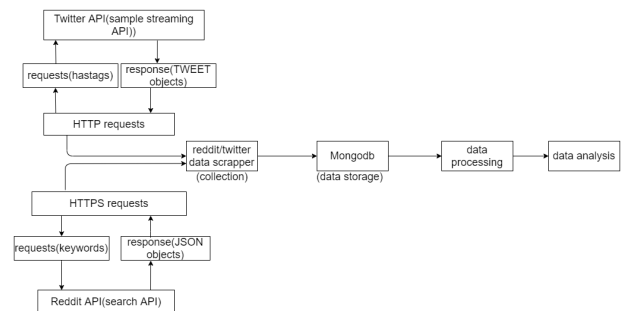


Figure 1: Data collection flow chart

3 DATA ANALYSIS

The data collected and stored in a data-base. We use mongoDB to store the data. The data is later cleaned before proceeding further.

We intend to perform a lexicon based analysis (aspect-based sentimental analysis / rule based approach) over the data using tools such as textBlob. Each data is analyzed for its polarity. We

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Woodstock '18, June 03–05, 2018, Woodstock, NY

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

<https://doi.org/10.1145/1122445.1122456>

shall categorize the positive tweets as pro-government and negative tweets as anti-government. Using the contrast between the counts for these two categories, we shall try to predict the results of the elections.

Furthermore, we can also analyze the contrast between the people's opinion on twitter and reddit using the same data.

4 DATA APPROXIMATION AND ESTIMATION

Twitter Search API processes 180 tweets/retweets per 15 min window per user. Assuming, we collect data throughout the day we would be able to collect 17k tweets data per day.

Reddit Search API processes 60 redds/subreddits per min per app. Assuming, we collect data throughout the day we would be able to collect around 86k redds/subreddits data per day.

Considering the above mentioned metrics, we intend to collect the data for around 10 days.

Therefore, combining the data collected from all our sources, we aim to collect approximately around 1 million count of data to make a solid prediction on the impact of racism on US elections.