

Impact of Abortion Rights on US Midterm Elections 2022

Ami Chauhan

State University of New York at
Binghamton University
Binghamton, New York, USA
achauha4@binghamton.edu

Kirtika Jawerilal

State University of New York at
Binghamton University
Binghamton, New York, USA
kjaweri1@binghamton.edu

Prarthna Mohanraj

State University of New York at
Binghamton University
Binghamton, New York, USA
pmohanr1@binghamton.edu

Suryavardhan Thummala

State University of New York at
Binghamton University
Binghamton, New York, USA
sthumma1@binghamton.edu

Ravi Teja Tadiboina

State University of New York at
Binghamton University
Binghamton, New York, USA
rtadibo1@binghamton.edu

ABSTRACT

Abortion has always been a controversy around the world. In 1973, the U.S. Supreme Court determined the verdict of the case of Roe vs. Wade. This decision proved to be a landmark one for the nation. The U.S. Supreme Court ruled that the decision whether to continue or terminate a pregnancy depends on the individual, not the government. However, in June 2022, the Supreme Court reversed the decision of Roe vs. Wade and declared that there is no longer a federal constitutional right to an abortion. In the light of these recent events, it seems that this issue will have a high impact on the midterm elections of the United States.

In this project, we intend to gain an overview of the general public and their thoughts on this particular issue and how it will affect the midterm elections held in November of this year. Social media is the best option to gather data for this. We will collect data from various forms of social media such as Twitter, Reddit, and YouTube to analyze the impact of abortion rights on the midterm elections being held.

KEYWORDS

Data sets	Abortion
Twitter	Elections
Reddit	Midterms
YouTube	US

ACM Reference format:

Ami Chauhan, Kirtika Jawerilal, Prarthna Mohanraj, Suryavardhan Thummala and Ravi Teja Tadiboina. 2022. IMPACT OF ABORTION RIGHTS ON US MIDTERM ELECTIONS 2022. In *Proceedings of ACM Woodstock conference (WOODSTOCK'18)*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/1234567890>

1 Introduction

The 2022 US Midterm Elections are done. All 435 seats in the U.S. House of Representatives and 35 out of 100 seats in the U.S. Senate were contested this year. The House of Representatives was won by the Republicans by a narrow margin while the Democrats maintain control of the Senate.

It has been a few months since Roe v Wade was overturned. Abortion rights has always been a “hot” topic here in the United States. But did abortion rights have an impact on the US Midterm elections held earlier this month? How much do voters think about abortion when voting between candidates?

We collected data during the month of November primarily. The three sources of data are Twitter, Reddit, and YouTube. Sentiment analysis was then performed [8] “to decipher the mood and emotions of general public and gather insightful information regarding the context.” We can gather a better understanding of the “various events and impact caused by it.”

2 Background and Related Work

2.1 Data Sources

Data used in this project was collected from the following social media platforms: Twitter, Reddit, and YouTube.

*Article Title Footnote needs to be captured as Title Note

*Author Footnote to be captured as Author Note

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WOODSTOCK'18, June 2018, El Paso, Texas USA

© 2018 Copyright held by the owner/author(s). 978-1-4503-0000-0/18/06...\$15.00

Twitter is the primary source of data. Twitter is one of the most popular forms of social media for political discussions. Teenagers to politicians use this form of social media to express their thoughts. Hence, a huge amount of data is available. Twitter has a number of APIs which can be used to retrieve and analyze Twitter data. We used Twitter's Sampled Stream API which gives us 1% of all tweets in real time. Twitter API consists of a set of programmatic endpoints which can be used to understand or build a conversation on Twitter. With the use of a persistent HTTP GET connection, the streaming endpoint delivers tweet objects.

Reddit was chosen as our second source of data. There are many active political discussion communities present on Reddit. We used Reddit's Search API to collect data. Data was collected if it was related to our keywords (i.e., 2022Elections, politics, abortion, Roe v Wade).

Lastly, data was collected from YouTube using the Google Client API. Data is in the form of comments that are posted for specific videos we have chosen that are related to how abortion is influencing the midterm elections.

3 Methodology

First, we collected data from Twitter, Reddit, and YouTube and stored the data in MongoDB. We then analyzed the data and created graphs which describe the data collected. Lastly, using TextBlob, we calculated the polarity and subjectivity of tweets and comments.

4 Data Acquisition

4.1 Twitter

Wikipedia describes Twitter[4] as an American microblogging and social networking service on which users post and interact with messages known as "tweets." Tweets were originally restricted to 140 characters. It was doubled to 280 for non-CJK languages in November 2017.

Twitter is a great data source due to the [1] "real-time nature of the content, and the ease in accessing and searching publicly available information. The reach and volume of data are also significant." Twitter can be a great source of data for researchers since it allows them to track, capture, and analyze users' activities and responses [2]. Twitter's streaming API is excellent for collecting data on a current event. All this put together makes Twitter an ideal source to gather people's opinion on the 2022 US Midterm Elections.

Twitter's sampled stream API [3] was used to collect data from Twitter. This method allows for real-time access to publicly available raw tweets. According to Twitter's API documentation: "The sampled stream endpoint delivers a roughly 1% random sample of publicly available Tweets in real-time. With it, you can identify and track trends, monitor general sentiment, monitor global events, and much more." You can connect one client per session, and can disconnect and reconnect no more than 50 times per 15 minute window. The code for this is implemented using python.

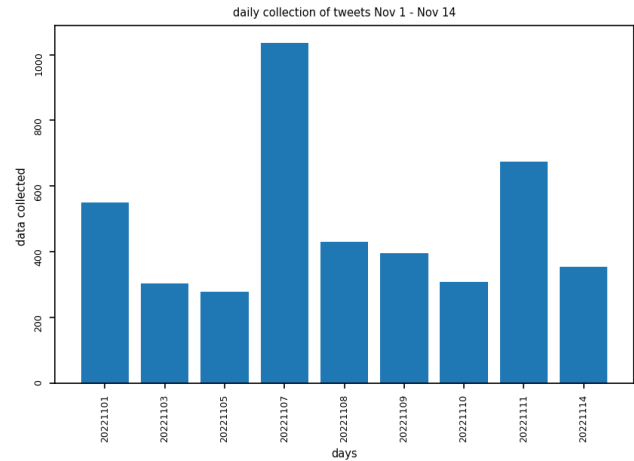


Figure 1: Twitter data collected from 11/01/22 to 11/14/22 shown daily

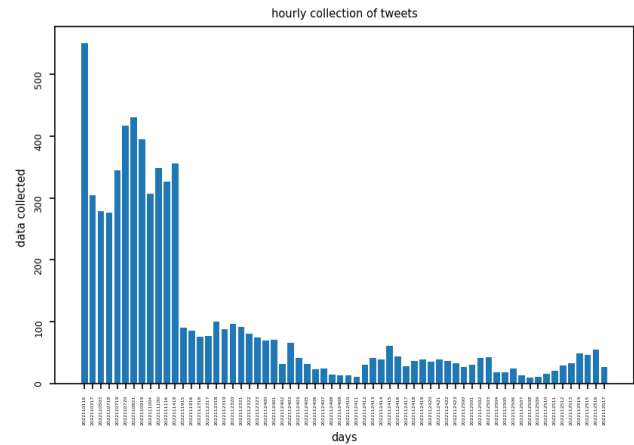


Figure 2: Twitter data collected from 11/01/22 to 11/25/22 shown hourly

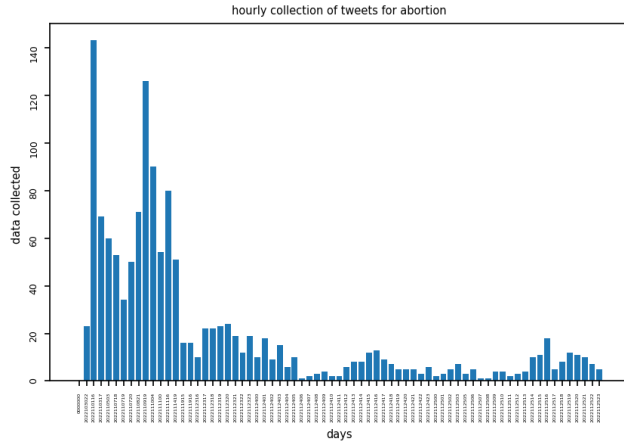


Figure 3: Twitter data collected from 11/01/22 to 11/25/22 for keyword 'abortion' shown hourly

4.2 Reddit

Reddit is an American social news aggregation, content rating, and discussion website. Redditors, or registered users, can “submit content to the site such as links, text posts, images, and videos, which are then voted up or down by other members. Posts are organized by subject into user-created boards called ‘communities’ or ‘subreddits’” [5] which cover topics from news to sports and celebrities.

The Reddit Search API [6] returns everything that matches our query. We need only the comments since sentiment analysis will be performed on the data. All the subreddits that are related to our query are returned by the Search API. This code was implemented in python. Figure 4 shows the number of Reddit comments collected from November 1 to November 25, 2022.

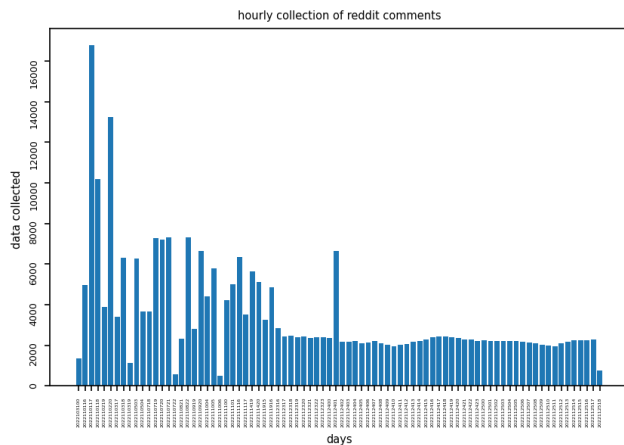


Figure 4: Reddit data collected from 11/01/22 to 11/25/22 shown hourly

r/politics is a subreddit on Reddit. This subreddit in particular is for news and discussion about U.S. politics. Currently, there are around 8.3 million members in this particular community. Figure 5 displays the submissions per day on the r/politics subreddit from November 4 to November 14, 2022.

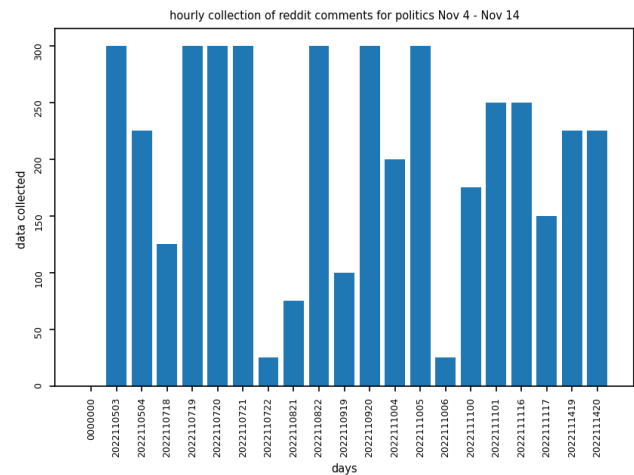


Figure 5: Reddit data collected from 11/04/22 to 11/14/22 for r/politics shown hourly

4.3 YouTube

YouTube is a social media and an online video sharing platform. There are more than 2.5 billion monthly users “who collectively watch more than one billion hours of videos each day”[7]. In addition, YouTube has a “unprecedented social impact, influencing popular culture, internet trends, and creating multimillionaire celebrities.”

Google Client API is used to collect comments from YouTube. Comments will be collected for only the specific videos that we have given. This code was implemented in python. Figure 6 shows the number of total comments for all our chosen videos from November 1 to November 18, 2022.

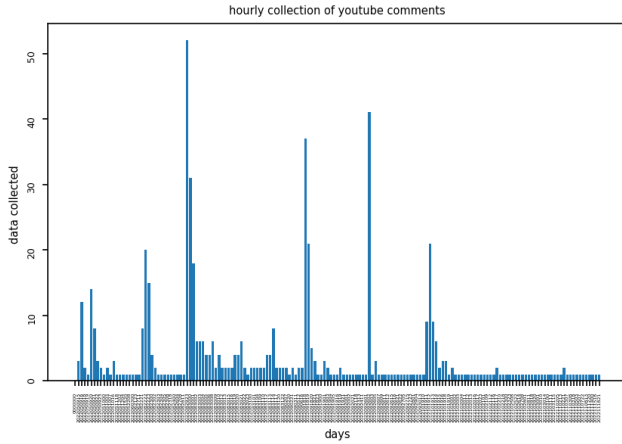


Figure 6: YouTube data collected from 11/01/22 to 11/18/22 shown hourly

5 Data Filtering

We collected data from Twitter, Reddit, and YouTube. YouTube did not contain as much data as the other two sources. Hence, we filtered all of our tweets and comments from Reddit and Twitter into two datasets: prolife and prochoice. We filtered the data so that it will be more specific and applied the same techniques to both Reddit and Twitter.

5.1 Relevant Datasets

We filtered all the tweets and comments collected from Twitter and Reddit into two datasets: prolife and prochoice. These datasets were created by using keywords related to prolife and prochoice.

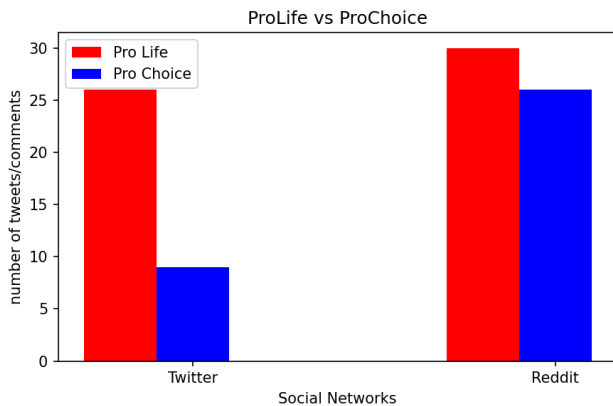


Figure 7: Representation of the two datasets (prolife and prochoice)

6 Sentiment Analysis

6.1 TextBlob

We decided to perform sentiment analysis on the tweets and comments collected using python libraries that are available at our disposal. We decided to use TextBlob which is a python library for Natural Language Processing [8]. Natural Language ToolKit (NLTK) "is a library which gives an easy access to a lot of lexical resources and allows users to work with categorization, classification, and many other tasks."

TextBlob returns the subjectivity and polarity of a sentence. "Subjectivity quantifies the amount of personal opinion and factual information contained in the text." High subjectivity value indicates that the text contains more personal opinion than factual information. Subjectivity lies between 0 and 1. On the other hand, polarity lies between -1 and 1. -1 represents a negative sentiment and 1 represents a positive sentiment.

Sentiment analysis was performed for tweets and comments from Twitter and Reddit that were filtered for abortion. All of the YouTube comments were analyzed. The subjectivity and polarity were determined for each piece of data. In addition, each tweet or comment was then analyzed to be either positive, negative, or neutral. The following graph represents the number of positive and negative tweets or comments for each data source.

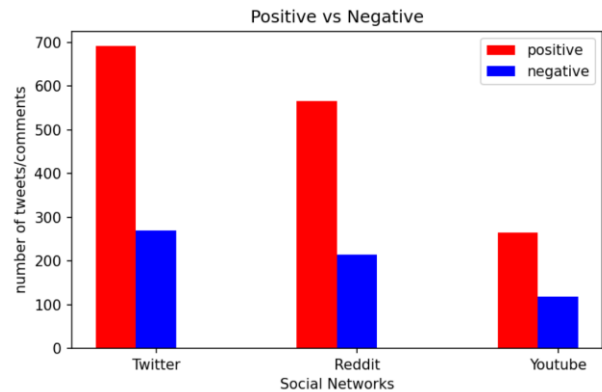


Figure 8: Representation of the number of positive and negative comments/tweets each data source

Number of Tweets/Comments					
	Source	Total	Positive	Negative	Neutral
0	Twitter	1397	691	269	437
1	Reddit	913	566	215	132
2	YouTube	574	264	118	192

Table 1: Representation of the number of positive, negative, and neutral tweets and comments

7 Limitations

7.1 Sarcasm and Context

Context plays a significant role in determining the meaning and intent of anything that has been communicated. By analyzing a few tweets manually, we found some tweets to be sarcastic or banter among friends. The contextual nature of the tweet has to be considered. Researchers have to take in the context and content of the tweet.

8 Results

Tweets and comments were analyzed using TextBlob. A total of 1397 tweets were filtered using the keyword 'abortion'. 49.4% of the tweets were found to be positive, 19.3% of the tweets were determined to be negative, and 31.3% of the tweets were found to be neutral. 914 comments were filtered from Reddit using the keyword abortion. 60.2% of the comments were found to be positive, 23.5% were found to be negative, and 14.4% were determined to be neutral. 45.9% of all the comments from YouTube were found to be positive, 20.6% were determined to be negative, and 33.4% of the comments was found to be neutral.

9 Conclusion

From all three of the data sources, we found that there was a higher percentage of positive tweets for abortion. Twitter, Reddit, and YouTube all have a very large number of users. Most of the users of these platforms are from the United States. Hence, we can assume that the opinions of the users from these social media platforms reflect the opinion of the majority of the people in this country.

ACKNOWLEDGMENTS

We would like to thank our Professor, Jeremy Blackburn, for his assistance with this project. We are very grateful for the feedback, support, and learning opportunities he has provided.

REFERENCES

- [1]2017. *Twitter as a Tool for Health Research: A Systematic Review*. Retrieved November 26, 2022 from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5308155/>
- [2]R. Rogers. 2014. Foreword: Debanalising Twitter: The Transformation of an Object of Study.

[3]2022. *Volume Streams*. Retrieved November 26, 2022 from <https://developer.twitter.com/en/docs/twitter-api/tweets/volume-streams/quick-start/sampled-stream>

[4]2022. *Twitter*. Retrieved November 26, 2022 from <https://en.wikipedia.org/wiki/Twitter>

[5]2022. *Reddit*. Retrieved November 26, 2022 from <https://en.wikipedia.org/wiki/Reddit>

[6]2022. *Reddit API Documentation*. Retrieved November 27, 2022 from <https://www.reddit.com/dev/api/>

[7]2022. *YouTube*. Retrieved November 27, 2022 from https://en.wikipedia.org/wiki/YouTube#Social_impact

[8]2020. *Sentiment Analysis using TextBlob*. Retrieved November 27, 2022 from <https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524>.