Unpacking the Opinions: A Deep Dive into Reddit's Communities on Pro-Life vs Pro-Choice. (1950 Words)

Jerin Easo Thomas
Spring 2023
Luddy School of Informatics,
Computing and Engineering
Indiana University Bloomington
Bloomington, Indiana
jerithom@iu.edu

INTRODUCTION

Pro-life and pro-choice are two diametrically opposed political and philosophical viewpoints on abortion. Pro-life supporters think that abortion should be prohibited and that the fetus has a right to life from conception. Pro-choice advocates, on the other hand, believe that women have the right to control their own bodies, including the option to have an abortion. Because these two topics symbolize broader social, cultural, and political tendencies and values, the public opinions about them are worth exploring. In addition, studying public sentiment on pro-life and pro-choice issues may help us better understand the underlying ideas and values that shape attitudes toward reproductive rights and women's health. Knowing the motivations and causes for public emotion on pro-life and pro-choice topics can also aid in the bridge-building process and develop better empathy.

A public health strategy put forth by *Sarah C. M. Roberts (2017)* is focused on increasing access to all reproductive health treatments, such as contraception, abortion services, and post-abortion care. It encourages patient-centred, evidence-based care and emphasizes the value of cooperation between public health officials, healthcare professionals, and community organizations[1]. Additionally, *Beth Reingold (2020)* offers a thorough analysis of the political and social factors that affect the formulation of anti-abortion policy, including the impact of conservative religious organizations, ideological conflict, and partisan politics[2]. *Heran Mane's (2022)* study analysed Twitter data and found patterns in the language used by people discussing the subject on the social media site. The author did this by using natural language processing techniques and sentiment analysis tools[3]. Using a qualitative analytical method, *Lauren K. Hunt (2022)* examined the information contained in #YouKnowMe posts on social media sites like Twitter and Instagram. The author places emphasis on issues related to having an abortion, including the challenges of finding safe and legal abortion services, the emotional and psychological toll of the procedure, and the requirement for support from family, friends, and medical professionals[4].

1. RESEARCH QUESTION

This technical publication's goal is to provide a complete examination of public opinion on pro-life and pro-choice issues. The study delves into two crucial questions:

1. What are the public's views toward the pro-life and pro-choice matters?

2. What are the most prominent supporting themes on either of these issues?

This study intends to elucidate the disparities in opinion among diverse populations, including those with neutral, positive, or negative views regarding these concerns, by analysing these questions. Furthermore, the study aims to uncover the primary ideas that support various points of view. Finally, by investigating the underlying feelings and circumstances that shape them, our research hopes to promote greater understanding between these two communities.

2. METHODS

In this study, a variety of methodologies were employed, of which some are as follows:

1. Python Reddit API Wrapper (PRAW): This package offers a straightforward interface for gaining access to Reddit's API. It makes it simple for users to engage with Reddit data, including posting new posts or comments, retrieving data on subreddits, comments, or people, and conducting content searches.

2. Valence Aware Dictionary and Sentiment Reasoner (VADER): This tool for sentiment analysis is lexicon and rule-based, and it was created to parse social media information. To ascertain the text's emotional tone, a list of terms and their corresponding sentiment scores are used. VADER is capable of handling emoticons, slang, and other informal language frequently used in social media. It can distinguish both positive and negative sentiment as well as neutral and mixed sentiment.

3. Latent Dirichlet Allocation (LDA): In topic modelling, LDA is a probabilistic model that is used to identify the underlying themes or subjects in a huge collection of documents. Using LDA, each document in the corpus will have its subjects and associated probabilities identified. Natural language processing and machine learning both frequently use LDA as a method.

4. Bidirectional Encoder Representations from Transformers (BERTopic): It is a topic modelling method and deep learning model that outperforms in a wide range of NLP applications, including sentiment analysis, question answering, and text classification. The resulting clusters serve as themes, with keywords picked for each cluster based on how effectively they reflect the issue.

3.1 DATA

This study's data came from two distinct subreddits: r/prochoice and r/prolife. The data is collected from January 2022 to March 2023 and is then divided into quarterly segments for sentiment analysis based on timeline. The mean upvote value of all posts and comments under the selected posts is evaluated so that posts and comments with more upvotes than the mean upvotes are deemed to be the top posts and comments, which would have the most information for which people are relating, and these top posts and comments are collected using the Python Reddit API Wrapper (PRAW) package along with some additional libraries such as Pandas and Natural Language Toolkit. To assure equality, a total of 1000 comments were collected in two batches: 500 comments from r/prochoice and 500 comments from r/prolife. Case insensitive keywords such as 'pro choice', 'prochoice', 'pro abortion', 'pro-abortion', 'pro life', 'prolife', 'anti abortion', and 'anti-abortion' are used to search for posts and comments in these subreddits because they capture the most relevant and informative data. Figure A illustrates all of the retrieved columns, with the most important being Comments, Date, Post Upvotes, and Comment Upvotes.

Fig A. Snapshot of collected Data

3.2 ANALYSIS

This research's analysis is carried out in four stages, which are detailed below.

Phase 1: Data gathering, analysis and cross validation.

Following the collection of data, this phase has concentrated on analyzing the fetched data. This includes data checks like ensuring that accurate and relevant data are being fetched from Reddit communities, ensuring that the fetched posts and comments have more upvotes than the calculated mean upvotes, cross-validating our fetched data with the data on Reddit by using the post URL, cleaning up redundant and unwanted data, ensuring that all column values are accurate,

ensuring that the fetched data is in readable format, and finally making sure the code was executed in two batches fetching not less than 500 comments from both the communities.

Phase 2: Sentiment Analysis and Visualization.

At this stage of the study, the Valence Aware Dictionary and the sEntiment (VADER) model, which is a pre-trained model for sentiment analysis, is used. Based on the input, which in our instance are the comments, the output of this model is a dictionary with four keys, Positive, Negative, Neutral, and Compound, and values representing the polarity score. Figure B shows an example of the model's output for a comment, indicating that the comment is neutral with the highest polarity score of, 0.598. According to the data, the majority of the comments are neutral; this could be because these topics are sensitive, and people are less likely to comment completely positively or negatively. As a result, I only evaluated positive and negative comments in this study to determine whether a comment is favorable or unfavorable.

Fig B. VADER output for a reddit comment

Phase 3: Timeline Sentiment Analysis.

The data, i.e. comments, are separated into four time frames in this step of analysis based on the date they were posted. During the year 2022, Q1 received comments from January 1 to March 31, Q2 received comments from April 1 to June 30, Q3 received comments from July 1 to September 30, and Q4 received comments from October 1 to December 31. Based on these timeframes, sentiment evaluations are performed using VADER, and the findings are compared quarter to quarter. The idea behind such an analysis is to understand how sentiments evolve over time.

Phase 4 : Topic Modelling

One key aspect of this research is identifying prominent themes noticed in the comments that may have an effect on people's views. To locate these themes, modelling with Latent Dirichlet Allocation (LDA) is first performed. Although this model does a good job of topic grouping associated terms, it fails to assign a single subject name to the resulting cluster. Figure C depicts three pro-choice positive comment clusters without the main topic classification. This is where Bidirectional Encoder Representations from Transformers (BERTopic) model provides an advantage by assigning the themes for each generated cluster of topics.

Fig C. Resultant clusters from LDA model

Based on all of the analysis performed throughout these four phases, one can obtain answers to questions like, What is the distribution of sentiment in both communities? What are the outcomes of comparing the attitudes of both communities? What are the most frequently used words in both communities comments? What are the most common themes discussed in both groups comments? Are there any signs of changes in the views of both communities over time?

3. RESULTS

This section includes detailed answers to each of the questions mentioned above, along with the results.

I. What is the distribution of sentiment in both communities?

According to the analysis, the bulk of comments in r/prochoice are negative (298 in total), with 202 favourable comments out of a total of 500. Similarly, the majority of comments are negative (296 in total) with 204 positive comments from the total of 500 in r/prolife, indicating that negative comments outnumber positive comments in both communities. Figure D and E shows the mentioned.
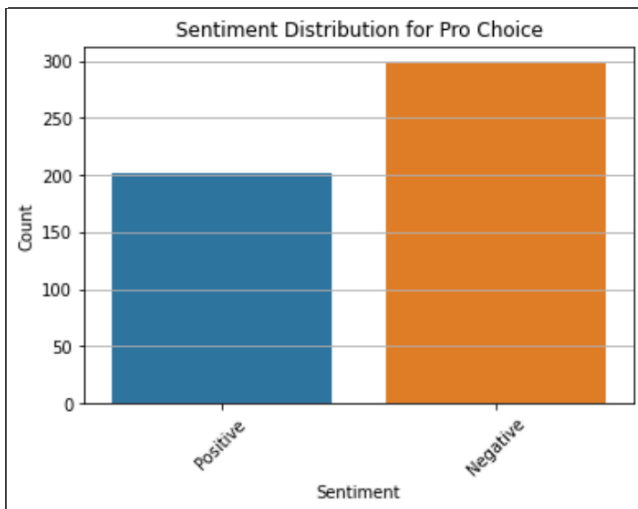

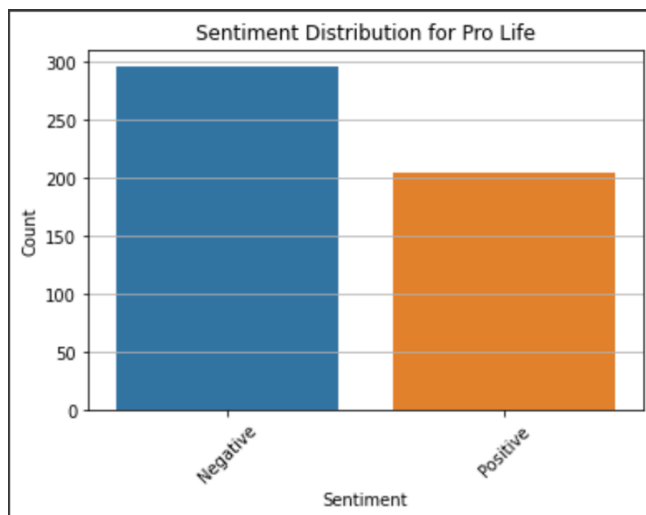Fig D. Sentiment distribution for pro-choice.

Fig E. Sentiment distribution for pro-life.

II. What are the outcomes of comparing the attitudes of both communities?

Based on the studies and comparisons, it is evident that the percentage of negative comments is higher in r/prochoice at 59.6% as compared to 59.2% in r/prolife, whilst percentage of positive comments is lower at 40.4% in r/prochoice as compared to 40.8% in r/prolife. Based on this, one can assume that the critical comments were written by members of the opposing community. Figure F and G depicts the same.
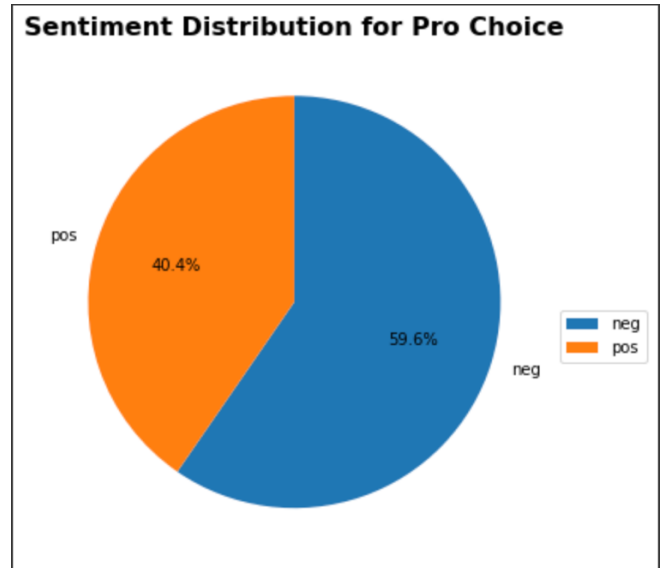

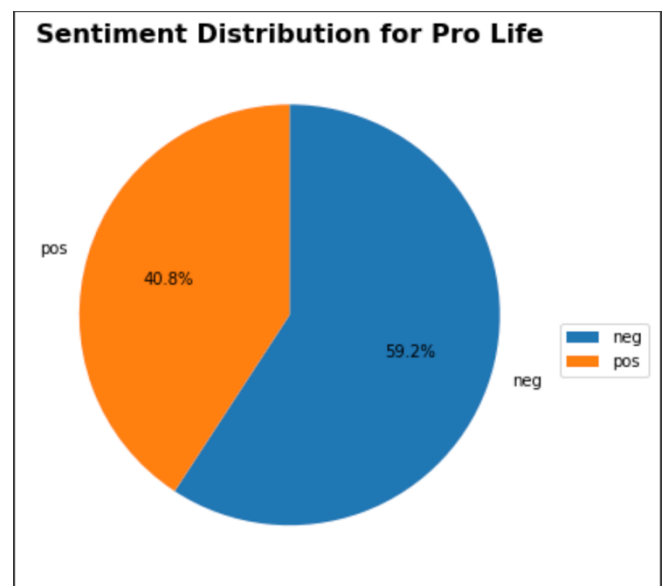Fig F. Sentiment distribution for pro-choice

Fig G. Sentiment distribution for pro-life

III. Are there any signs of changes in the views of both communities over time?

Based on sentiment analysis throughout time, one can notice a cyclic repetition of feelings in r/prochice comments every two quarters, which could be attributed to some external occurrences. Yet, the comments from r/prolife initially show more negative remarks than positive comments in the first quarter, but in later quarters, positive comments (300+) outnumber negative comments. Figure H and I shows the aforementioned.
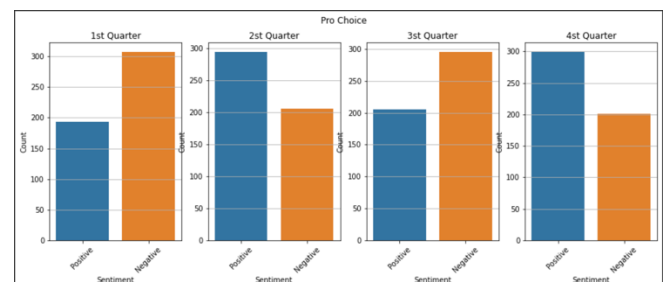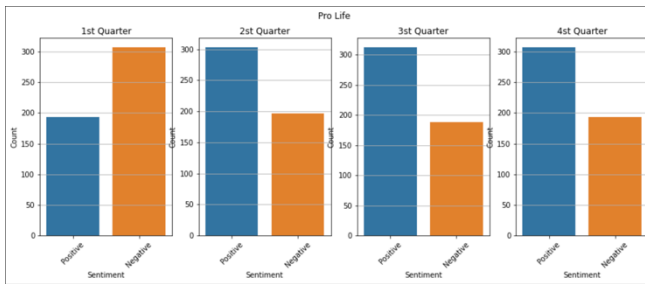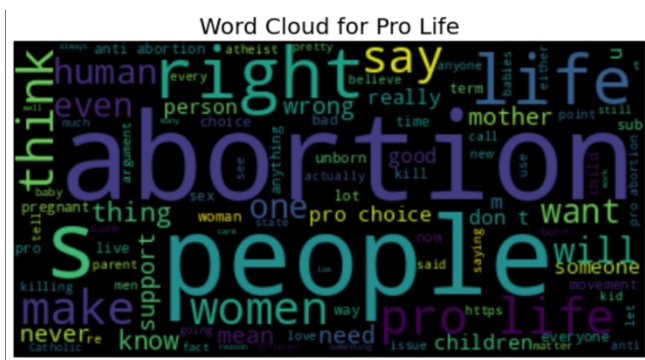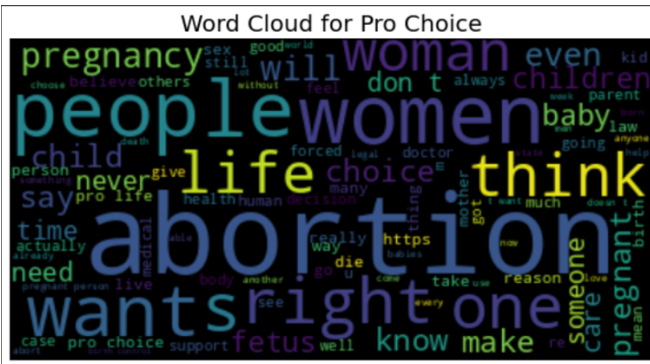

Fig H. Sentiment distribution for pro-choice over four quarters

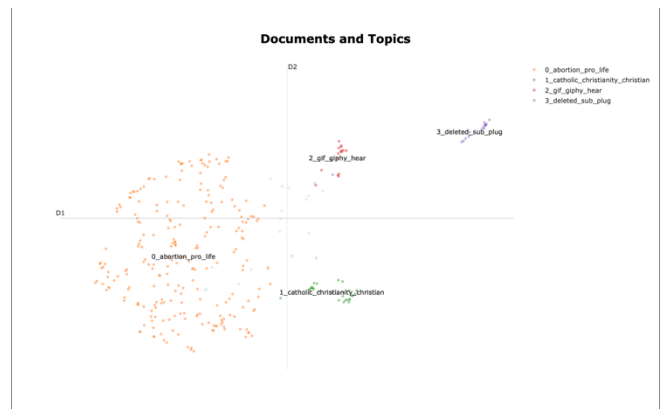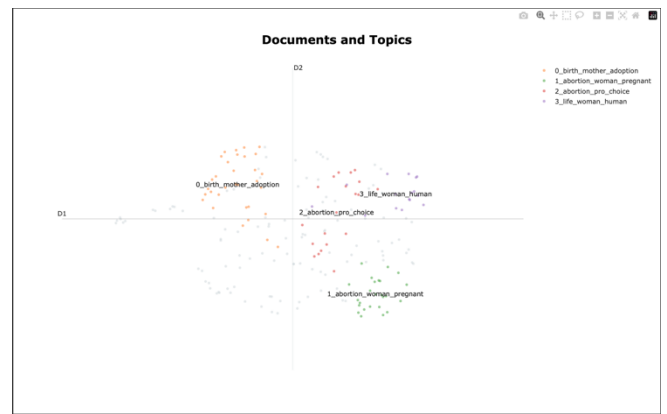*Fig I. Sentiment distribution for pro-life over four quarters*

IV.   What are the most frequently used words in both communities comments?

By building word clouds for both the communities i.e. r/prolife and r/prochoice, the most often used words in the comments in r/prochoice are 'abortion', 'women', 'people', and 'want' with values 1.0, 0.49, 0.48, 0.41 respectively, while in r/prolife the most frequently used words in the comments are 'abortion', 'people', 'right', and 'life' with values 1.0, 0.81, 0.53, 0.48 respectively. The same are seen in Figure J and K.

Fig J. Word cloud for pro-choice

Fig K. Word cloud for pro-life

V.   What are the most common themes discussed in both groups comments?

BERTopic topic modeling was applied in four sections: pro-choice positive remarks, pro-choice negative comments, pro-life positive comments, and pro-life negative comments and examples of the resulting themes were 'human women life', 'pregnant women abortion', and 'abortion pro life'. By hovering over and reading the comments of each cluster in these sections, it is possible to determine that comments from r/prolife support the right to life while comments from r/prochoice support the freedom to choose.

Fig L. Topics in pro-choice positive comments

Fig M. Topics in pro-choice negative comments

4. CONCLUSION

Finally, this research article investigated the attitude, trend, and importance of abortion-related discussions in the Reddit communities of r/prochoice and r/prolife. The data indicated that both communities mostly expressed unfavourable attitudes towards anti-abortion and pro-abortion groups, with r/prolife showing an increase in positive comments over time. Furthermore, r/prochoice supports women's rights, whereas r/prolife prioritized the right to life based on the results from the BERTopic model. The data confirmed that members of r/prolife are generally opposed to abortion, whereas members of r/prochoice are pro-abortion. Overall, the study emphasizes the disparities in the two communities opinions and interests on the sensitive topic of abortion.

Limitations:

Given that this is a delicate subject, both communities may be reluctant to provide totally positive or negative feedback. The comments could potentially be skewed due to external events or personal experiences. Although, the current model captures the core theme, considerably more powerful models might be employed to determine the actual reason for clustering.

5. REFERENCES

[1] Sarah C. M. Roberts, Liza Fuentes, Nancy F. Berglas, and Amanda J. Dennis. A 21st-Century Public Health Approach to Abortion. Retrieved from Am J Public Health. 2017 December; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5678382/

[2] Beth Reingold, Rebecca J. Kreitzer, and Michele L. Swers. Anti-abortion Policymaking and Women's Representation. Retrieved March 4, 2020. https://journals.sagepub.com/doi/full/10.1177/1065912920903381

[3] Heran Mane. Examination of the Public's Reaction on Twitter to the Over-Turning of Roe v Wade and Abortion Bans. Retrieved from National Library of Medicine. 2022 December; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9777967/

[4] Lauren K. Hunt, Carly M. Goldstein, Camille L. Garnsey, Katherine W. Bogen and Lindsay M. Orchowski. Examining Public Sentiment Surrounding Abortion: A Qualitative Analysis of #YouKnowMe. Retrieved January 20, 2022. https://www.tandfonline.com/doi/abs/10.1080/23293691.2021.2016161

[5] Linda J Beckman, Abortion in the United States: The continuing controversy. Retrieved February 1, 2017. https://journals.sagepub.com/doi/10.1177/0959353516685345

[6] PRAW- Python Reddit API Wrapper. Retrieved from https://www.geeksforgeeks.org/python-praw-python-reddit-api-wrapper/

[7] Sentiment Analyses using VADER. Retrieved from https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664

[8] Interactive Topic Modelling with BERTopic. Retrieved from https://towardsdatascience.com/interactive-topic-modeling-with-bertopic-1ea55e7d73d8