

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/369529057>

Frequency and Sentiment Analysis of Discussions on Reddit Religious Communities

Research · March 2022

DOI: 10.13140/RG.2.2.10610.99523

CITATIONS

0

READS

4

1 author:



[Hadi Abbas](#)

Indiana University Bloomington

9 PUBLICATIONS 0 CITATIONS

SEE PROFILE

Frequency and Sentiment Analysis of Discussions on Reddit Religious Communities (1435 words)

Hadi Abbas

Luddy School of Informatics, Computing, and Engineering, Indiana University Bloomington

ILS Z-639: Social Media Mining

Ali Ghazinejad

March 27, 2022

Abstract

This paper analyzes the discussions on five major religious communities on the social media platform Reddit. The study seeks to explore the topics and themes discussed in the Islam, Christianity, Judaism, Hinduism, and Atheism subreddit communities by scraping user comment data from the last year. Specific research interests for this study include a comparison amongst the communities with regards to frequently discussed words and overall discussion sentiment.

Keywords: social media, religion, natural language processing, sentiment analysis, atheism

Frequency and Sentiment Analysis of Discussions on Reddit Religious Communities

Introduction

As one of the leading social media platforms today, Reddit has provided users from all over the world to join communities based on their interests and the groups they identify with. This includes politics, technology, entertainment, sports, news, and religion. The latter is particularly interesting given how heated and uncivil discussions regarding beliefs sometimes can get. Furthermore, there is an ongoing widespread shift away from traditional religious beliefs in the postmodern world as an increased number of people, including many younger people, are embracing atheistic or agnostic ideals. On the other hand, there remain billions of people globally who continue to look to religion for spiritual guidance and a sense of purpose. In a recent study titled *Distant Reading of Religious Online Communities: A Case Study for Three Religious Forums on Reddit*, Thomas Schmidt, Florian Kaindl, and Christian Wolff at the University of Regensburg, Germany analyzed three religious subreddits, r/Islam, r/Christianity, and r/Occult, via computational text analytics (2020). This study analyzed word frequencies, bigram frequencies, and collocations in the three communities to determine what posts and comments were closely associated with keywords like “god”. In another study, Evelina Lundmark of Uppsala University and Stephen Ledrew of the Memorial University of Newfoundland analyzed user-generated content on r/atheism to determine how users arrived at their atheistic beliefs (2019). Similar studies have also been conducted on other social media platforms, such as *Sentiment Analysis on YouTube: A Brief Survey* published in 2015 by Muhammad Zubair Asghar, Shakeel Ahmad, Afsana Marwat, and Fazal Masud Kund; this partly dealt with the analysis of the civility of religious comments on YouTube videos. This study seeks

to expand upon some of these previous works by comparing five unique religious communities through Reddit comments.

Research Questions

What are users on the religious subreddits r/Islam, r/Christianity, r/Judaism, r/Hinduism, and r/Atheism posting and commenting about? What are some key similarities and differences across the different communities?

Methods

Data

The top post from each of the five subreddits (r/Islam, r/Christianity, r/Judaism, r/Hinduism, and r/Atheism) between March 2021 and March 2022 represents the initial starting point of data collection for this study. These five communities, in particular, were selected to include the primary subreddits of the two most practiced religions in the world, Christianity and Islam, that of the other of the three traditionally Abrahamic monotheistic religions, Judaism, and the third-most widely followed religion in the world, Hinduism. Lastly, the largest subreddit dedicated to atheism was also included to provide a largely antithetical perspective to the concept of religion than endorsed by members of the former four groups. The one-year time window was selected to avoid recency bias and provide a reasonable, equal time frame across the communities. For example, The corresponding top-level comments from each of these five posts were then collected using the Python Reddit API Wrapper (PRAW) and web scraping tools. For each unique subreddit, the comments were stored in data frames via the pandas library. After using data cleansing to drop deleted and removed comments, the total number of remaining comments per community was as follows: 119 for Islam, 78 for Christianity, 38 for Judaism, 86

for Hinduism, and 86 for Atheism. This results in a total of 407 comments that were analyzed in this study, with an average of 81.4 per community.

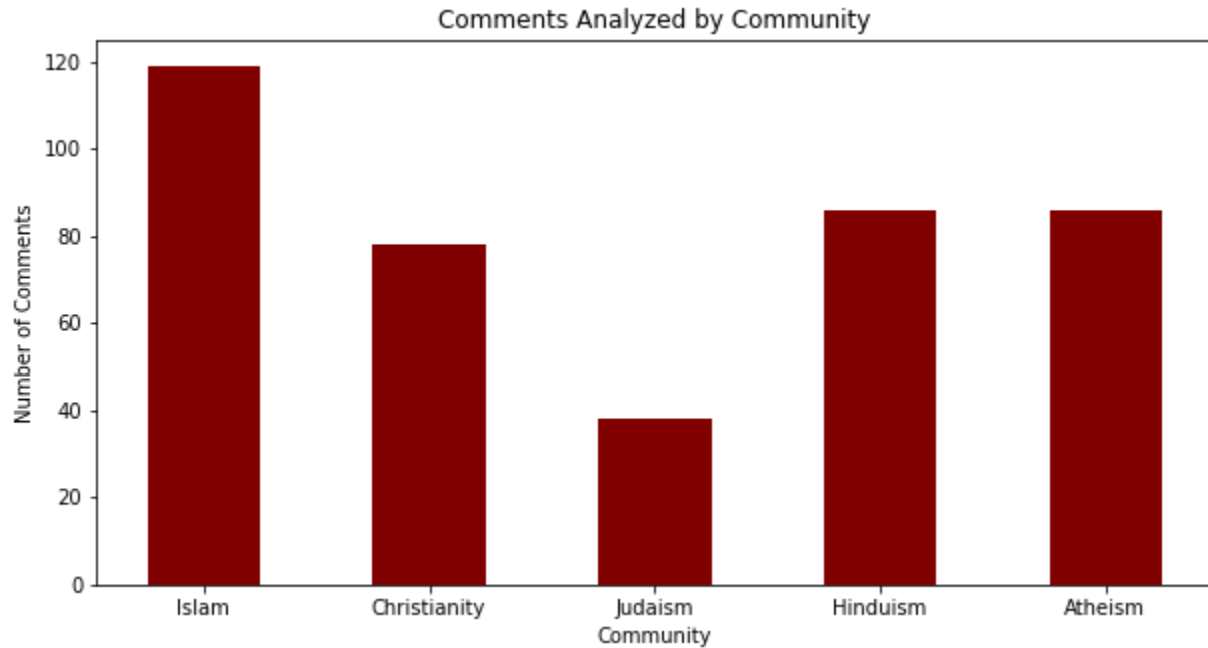


Figure 1

Analysis

The comments were firstly compared via word frequency counts. This process required additional libraries including Natural Language Toolkit (nltk) and Collections. The data were converted into entirely lowercase values and further standardized by removing punctuation items such as apostrophes. From here, the most frequent words and corresponding word counts for all five communities were computed and then compared via data visualizations.

Next, the data was analyzed for sentiment analysis using the vaderSentiment library. In this stage of the study, the lines in the subreddit-specific data frame were concatenated into a single string value for accessibility. These strings were then individually passed into a manually created function that summarizes the sentiment of the overall discussion using the

polarity_scores method of the Sentiment Intensity Analyzer from the aforementioned vaderSentiment library. This function also determines the percentages of the three different sentiment categories (positive, negative, and neutral).

Results

Outcome 1: Word Frequencies

The most common words and corresponding counts in the comments sourced from the Islam subreddit are as follows: people (26), israel (23), like (14), dont (12), and even (11). For Christianity, the counts of the most frequent words are god (20), pray (16), bless (13), children (12), and like (12); for Judaism, they are people (8), like (7), dont (6), antisemitism (5), and day (5). Looks (28), game (22), amazing (17), would (17), good (13), and like (13) make up the most frequently used words from the comment section in the top post in r/Hinduism. For Atheism, the most common words are chicken (22), bk (18), burger (16), king (15), sandwich (15), and food (15).

Outcome 2: Sentiment Analysis

As determined by the polarity index described earlier, the overall discussions in r/Christianity, r/Hinduism, and r/Atheism were rated as positive while the discussion in r/Islam and r/Judaism was rated as negative. The highest overall negativity score was 19.6 % in r/Islam and the highest overall positivity score was 31.8% in r/Hinduism. Somewhat unsurprisingly, these two subreddits swapped places when it comes to the lowest overall negativity score (4.0 % in r/Hinduism) and the lowest overall positivity score (13.0 % in r/Islam). However, the

sentiment was overwhelmingly neutral across all of the observed communities. All of the analyzed discussions had a neutral rate of above 62.9 %.

Conclusion and Limitations

This study explored five distinct religious communities on the social media platform Reddit and utilized data science and natural language processing tools to compare and contrast frequently used words and discussion sentiments between them. The top-level comments from the top post over the last year were used as the dataset for the research conducted in this study. The most frequent words in the three Abrahamic communities of Islam, Christianity, and Judaism contained many words one might expect when it comes to discussing religion like “god”, “pray”, and “bless”. The topics discussed in r/Islam and r/Judaism were both related to the Israel-Palestine conflict and therefore the words “people”, “israel”, “anti”, “antisemitism”, “palestine”, “hate”, and “terrorists” are among the frequently mentioned in these two subreddits. On the other hand, the post in r/Hinduism was more motivational and therefore had frequent words that one might not often expect to be mentioned when discussing religion. Lastly, r/Athiesm also has frequent words that can be misunderstood without the context that the parent post was about Burger King donating money to pro-LGBTQ organizations in response to Chick-fil-A.

The second aspect of the analysis focused on overall discussion sentiment with natural language processing. Here, it was determined that the Islamic community had both the most negative and least positive sentiment while Hinduism had both the most positive and least negative. In addition to Islam, Judaism was the only other community to have an overall negative sentiment. All five subreddits were found to have a significant amount of neutral sentiment given

the constraints set forth by the researcher for this study. Once again, the post topic for the discussion played a major role as r/Islam and r/Judaism were discussing heated topics compared to the other three communities.

This study has a fair amount of limitations and thus, many areas for improvement that could allow for further research. Focusing on more than one post from each community would likely allow for more variance and balance in regards to what topics are being discussed. Rather than attempting to cover five different communities, a more detailed study could be conducted focusing on only two subreddits, or even just a single one. Furthermore, the word frequency comparison and sentiment analysis used for this study are just two of several methodologies by which the Reddit comment data could be compared and contrasted.

References

- Lundmark, E., & LeDrew, S. (2019). Unorganized atheism and the secular movement: reddit as a site for studying 'lived atheism.' *Social Compass*, 66(1), 112–129.
<https://doi.org/10.1177/0037768618816096>
- Schmidt, T., Kaindl, F., & Wolff, C. (2020). Distant Reading of Religious Online Communities: A Case Study for Three Religious Forums on Reddit. *DHN*, 157–172.
- Zubair Asghar, M., Ahmad, S., Marwat, A., & Masud Kundi, F. (2015). Sentiment Analysis on YouTube: A Brief Survey. *MAGNT Research Report*.