Sentiment Analysis of Reddit & YouTube

A Dataset Measurement & Analysis Report

Venkata Purna Vishnu Vardhan Kolleboyena
Computer Science
Binghamton University
Binghamton, New York, USA
kvenkatapurn@binghamton.edu

Avinash Kunamneni
Computer Science
Binghamton University
Binghamton, New York, USA
akunamneni@binghamton.edu

Shashank Reddy Karnati
Computer Science
Binghamton University
Binghamton, New York, USA
skarnati@binghamton.edu

ABSTRACT

In the second phase of our project, our primary objective was to analyse and address negativity and toxic behaviour on social media platforms, specifically Reddit and YouTube. Leveraging sentiment analysis and modern hate speech detection APIs, we delved into the emotional and toxicity levels within these online communities. Through extensive data visualization, we compared Reddit and YouTube, offering insights into the prevalence of toxic content and the impact of diverse opinions on these platforms. The results of our analysis, depicted in meticulously crafted plots, serve as a foundation for further investigation into the interplay between sentiment and hate speech. Our research endeavours to contribute to fostering a more positive and respectful online environment for all users.

KEYWORDS

RedditAPI, YouTubeAPI, Java, Http Client, Python, Matplotlib, PostgreSQL, Text Blob, Data Analysis, ModernHateSpeech API, Data Collection, Data Visualisation,

ACM Reference format:

Venkata Purna Vishnu Vardhan Kolleboyena, Avinash Kunamneni and Shashank Reddy Karnati. 2018. Sentiment Analysis of Reddit and YouTube: A Data collection and analysis pipeline. In *Proceedings of ACM Woodstock conference (WOODSTOCK'18). ACM, New York, NY, USA, 2 pages.* https://doi.org/10.1145/1234567890

1 Introduction

In the second phase of our project, we are committed to tackling the growing issue of negativity and toxic behaviour within the realms of social media platforms, particularly Reddit and YouTube. Our strategy involves a two-fold approach, employing both sentiment analysis and the ModernHateSpeech API to delve deeper into the intricacies of these online communities.

Our project began by harnessing the power of TextBlob to conduct sentiment analysis on the extensive corpus of user-generated comments. This analysis provided valuable insights into the emotional and attitudinal underpinnings of these interactions. By categorizing sentiments as positive, negative, or neutral, we uncovered patterns and trends in user behavior, offering insightful perspectives.

In parallel, we leveraged the ModernHateSpeech API to detect and quantify the presence of toxic content within these platforms. This specialized tool aided us in identifying and categorizing harmful content, encompassing hate speech, insults, and offensive language. Our goal was to contribute to the creation of a more respectful and positive online environment.

By employing these two distinct yet complementary approaches, we gained a comprehensive understanding of the emotional landscape and the extent of toxicity within these online communities. Through our research and the presentation of data visualizations, we shed light on the impact of differing opinions, providing insights that pave the way for a more constructive and harmonious online space for all users.

2. System Design for Data Analysis

As we are collecting the data continuously, we are simultaneously running a python script which processes the data records in batches of 1000 records at a time to evaluate and update the sentiment score and hate speech detection.

We are retrieving the records that are being collected in batches and then at first, we are performing the sentiment analysis. Subsequently, for the same data we are performing the API call to the modern hate speech API which will return us the results. We are mapping those two fields back into the data base against those data records. You can find a simple overview of what we are going to perform in the below figure 1. The figure assumes entire Project 1 work as a Data collection System being in place.

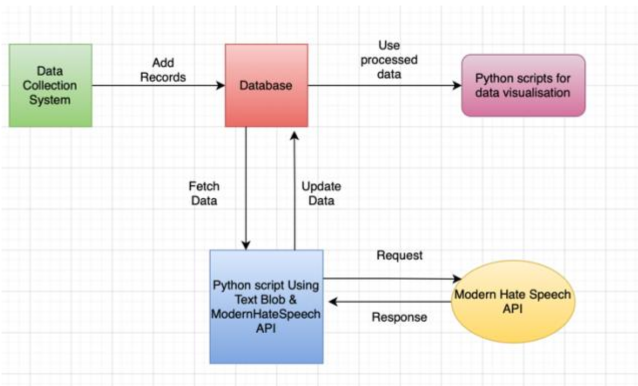

Figure 1

We are effectively handling errors from the ModernHateSpeech API by assigning a "api-error" flag to the data records in the database in the event of a failure. We conduct retries on all those records marked with an "error" flag every weekend to ensure their processing as well. You can find a detailed description of how these are being handled in the section 4.2.

3. Background & Motivation

In the vast realm of online communication, the intricate dance between positive discourse and toxicity has become a pressing concern. This project is motivated by the recognition that social media platforms serve as influential conduits for shaping public opinion and fostering diverse conversations. As we navigate this complex digital landscape, understanding and mitigating issues of hate speech and varying sentiments have become paramount for cultivating healthy online communities.

Our motivation is grounded in the wealth of existing scientific literature that delves into the various aspects of user-generated content. By referencing and finding inspiration from papers like "Understanding the Behaviors of Toxic Accounts on Reddit" and "Analysing Sentiments for YouTube Comments using Machine Learning" cited in our references, we aim to actively contribute to the ongoing conversation about forging safer and more inclusive digital spaces. The literature review, neatly encapsulated in our reference section, serves as a nod to the insightful work of researchers who have paved the way. It shines a light on the diverse knowledge that guides our approach, emphasizing the importance of translating theoretical insights from these papers into practical solutions for real-world challenges.

The exploration of sentiment analysis and hate speech detection emerges as a natural progression from the foundational work of researchers who have grappled with the complexities of online communication. Through a systematic analysis of existing literature, we aim to bridge the gap between theoretical understanding and actionable strategies.

As we embark on this research journey, the synthesis of theoretical frameworks and real-world applications forms the backbone of our approach. By aligning our motivations with the wealth of existing knowledge, we aspire to offer nuanced perspectives and practical solutions that contribute to a more positive and respectful digital environment for all users.

4. Data Description

Now that we have started to run our data analysis code and we have obtained enough data for analysis, below are the descriptions of our data that has been collected so far. **From here on, we are going to treat “r/politics” subreddit as a separate data source in all the tables and plots.** The data shown below is inclusive of both English and non-English comments & the count is subject to change (data included in the report is taken on Nov-29,2023).

Data Source	Comments Count
Youtube	823764
Reddit	1197864
Politics	575088

Below is the data of the above 3 data sources in a more detailed view.

Youtube:

We have collected comments from the various English singer’s music video’s as shown below.

Video Id	Comments count
VuNIsY6JdUw	346218
WcIcVapfqXw	169300
yOuqn4w1ozA	137076
lVkKLf4DCn8	59196
ixkoVwKQaJg	41675
suAR1PYFNYA	24486
rYEDA3JcQqw	22611
OiC1rgCPmUQ	18403
XzOvgu3GPwY	4899

Reddit:
The reddit data source contains the list of below subreddits and politics subreddit is treated as different data source altogether.

Subreddit	Comments Count
r/wordlnews	195785
r/gaming	195211
r/movies	192918
r/soccer	174720
r/news	166932
r/conspiracy	152747
r/science	46750
r/gameofthrones	44783
r/breakingbad	18158
r/TrueReddit	7171
r/offbeat	2773
r/announcements	25

Politics:
The count shown below is right from the start of the data collection system. And the data between Nov-1 to Nov-14 has been plotted separately in the Figures 4,5.

Subreddit	Comments Count
r/politics	575088

4.1 Data Visualisation:

The data has that has been shown so far for various data sources has been plotted down in form of line graph and bar graphs as well. We have plotted down the count of comments per each day over last 30 days.

Comments per each for last 30 days:

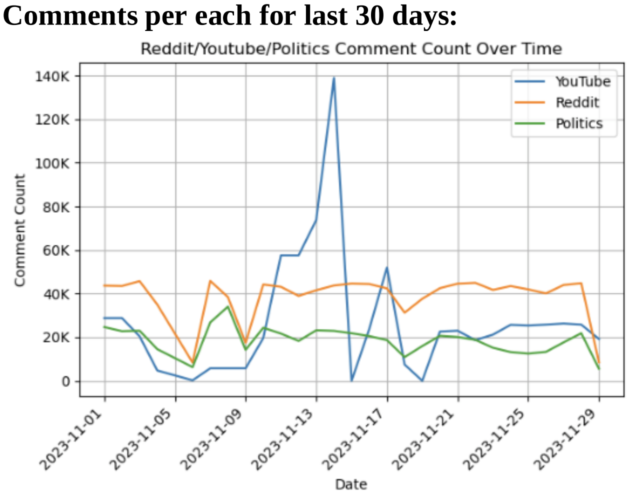

Figure 2

Total Count of comments so far:

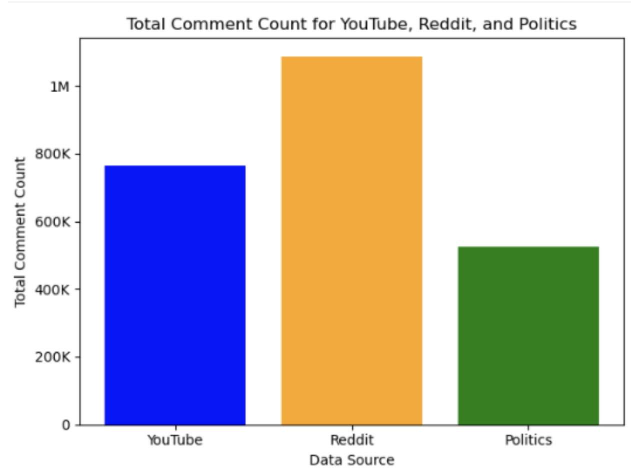

Figure 3

The “r/politics” data that has been collected from Nov-1 to Nov-14 has been plotted below.

No of Submissions binned daily (Nov1-14):

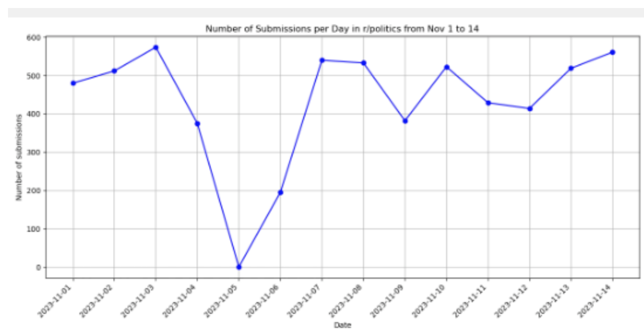

Figure 4

No of comments binned hourly (Nov1-14):

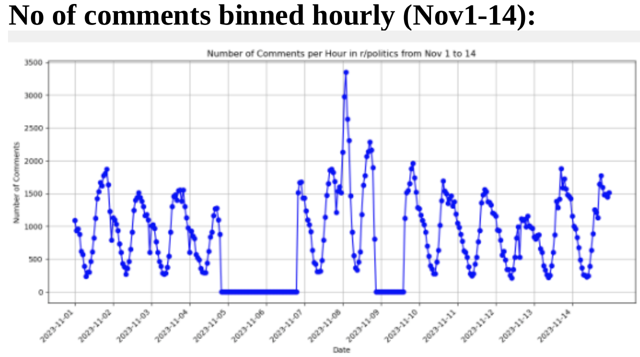

Figure 5

Similarly, here’s a bar graph that is plotted to give us an overview count of comments collected in each subreddit excluding politics.

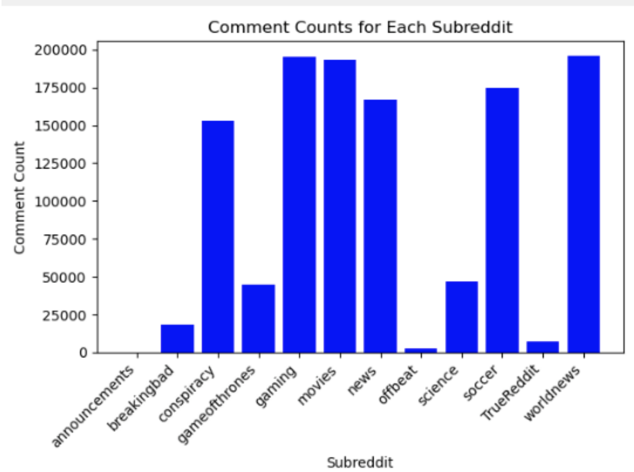

Figure 6

Here's our cumulative plot of the projected estimate vs the actual estimate. Below are the plots and the analysis for why there's a difference in the projected estimates vs actuals has been explained in the section 5.

Projected at the time of Project 1:

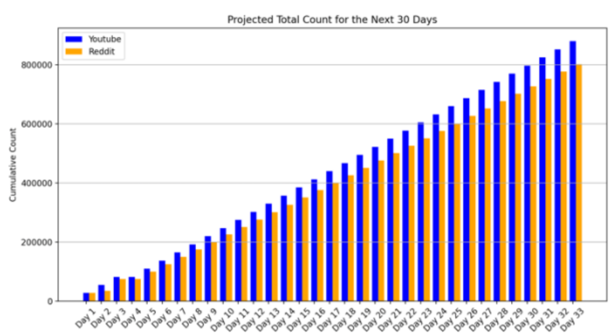

Figure 7

Actuals at the time of Project 2:

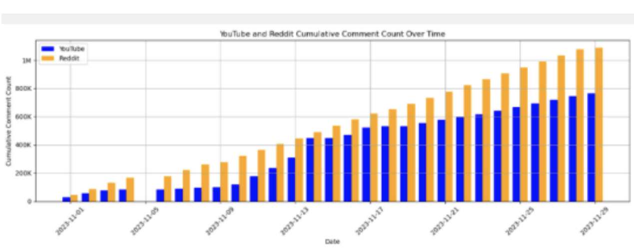

Figure 8

As we delve into the realm of sentiment analysis and hate speech detection, a crucial step in our methodology involves meticulously cleaning the data to distinguish between English and non-English comments. Recognizing the efficacy of TextBlob and the ModernHateSpeech API specifically for English language comments, we employ the "langdetect" Python library to discern the language of each comment. Only those identified as English undergo further analysis, ensuring a targeted approach to sentiment analysis and hate speech detection.

Hence, the graphs plotted for both the sentiment analysis and hate speech using ModernHateSpeech API data uses the following count of comments that are in English language. We have excluded the non-English language comments for Sentiment & hate speech analysis.

Data Source	Count of English Comments
Youtube	441715
Reddit	1119970
Politics	548188

In our process, sentiment analysis is conducted using the TextBlob library, which assigns a sentiment polarity to each English comment—categorized as positive, neutral, or negative. Simultaneously, we utilize the ModernHateSpeech API, applying a default threshold of 0.9 as per the API documentation. Comments exceeding this threshold are classified as hateful, while those below are labelled as not hateful. The language information, sentiment polarity, and hate speech classification are meticulously stored for each data record. To streamline processing, we operate in batches of 1000 records at a time.

In instances where an error arises during processing, we adopt a robust error-handling strategy by marking the corresponding hate value field with "api-error" for future retries. Regular reprocessing of these flagged records occurs every weekend to ensure comprehensive analysis. For comments identified as non-English, we categorize them as "not applicable" in our dataset. Consequently, the ensuing graphs and data discussions pertain exclusively to English comments, with non-English comments distinctly separated from the visualizations presented below. This meticulous approach ensures a focused and accurate analysis of sentiment and hate speech within the realm of English-language comments.

4.2 Error Handling:

We are mapping the hate-value field as "api-error" in case of an error from the ModernHateSpeechAPI and there has been numerous outages of that API over the last 30 days and the data can be seen below.

```
youtube=# select count(*) from comments where hatevalue='api-error';
 count
-------
 22571
(1 row)

youtube=#
```

```
politics=# select count(*) from reddit_comments where hatevalue='api-error';
 count
-------
 25969
(1 row)

politics=# 
```

```
postgres=# \c reddit
You are now connected to database "reddit" as user "postgres".
reddit=# select count(*) from reddit_comments where hatevalue='api-error';
 count
-------
 73367
(1 row)

reddit=# 
```

We have written a python script to run the failed records over every weekend and are trying to minimise the loss of data.

4.3 Plots of Sentiment Analysis:

Below are the plots that we have obtained by running our python scripts to classify the English comments as "Positive","Neutral"&"Negative". You can find the plots for all the three data sources. The plots below contain the data obtained from the overall list of English comments in all three data sources.

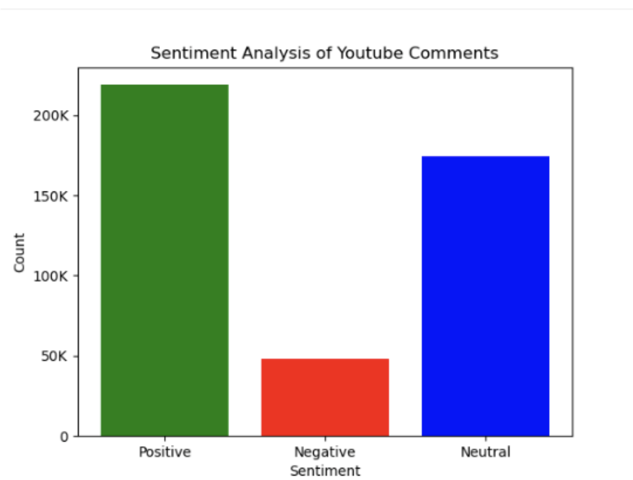

Figure 9(a)

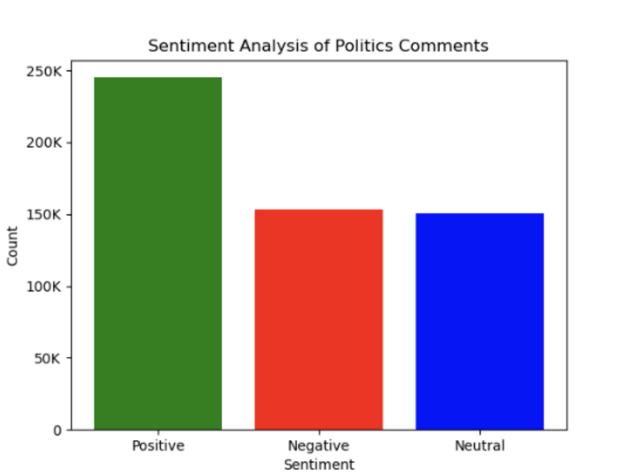

Figure 9(b)

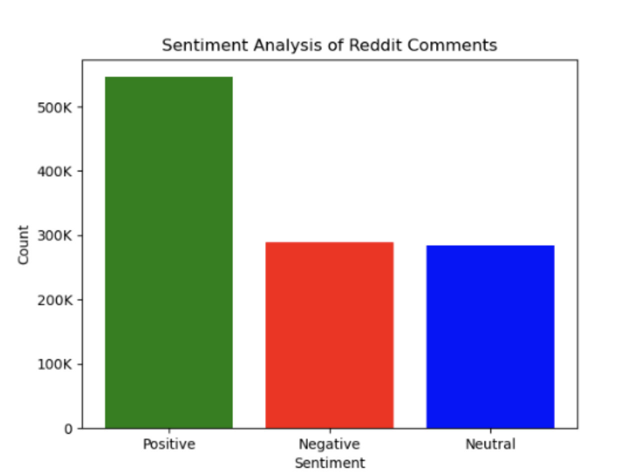

Figure 9(c)

4.4 Plots of HateSpeech using ModernHateSpeech API data:

Below are the plots that we have obtained after we have processed our English comments using the ModernHateSpeech API. We have classified them as "hateful" and "not hateful". You can find the plots for all the three data sources. The plots below contain the data obtained from the overall list of English comments in all three data sources.

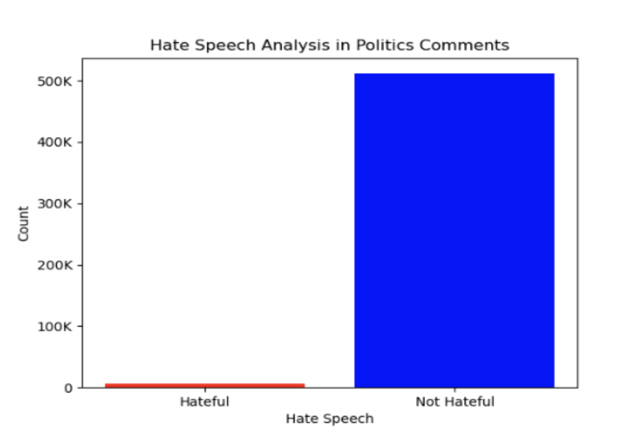

Figure 10(a)

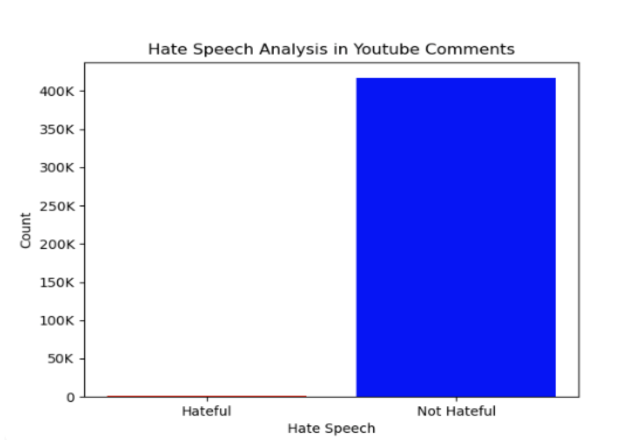

Figure 10(b)

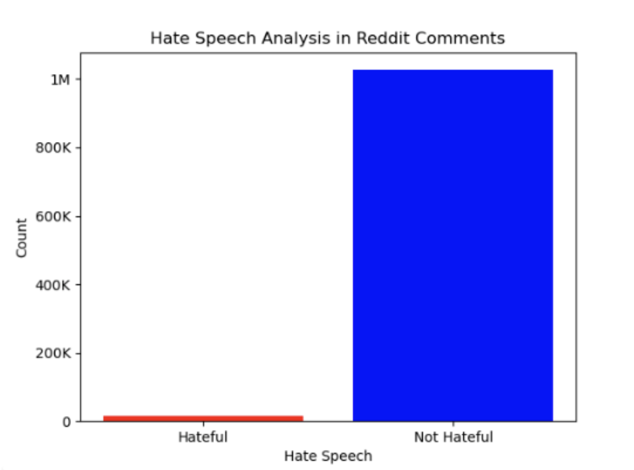

Figure 10(c)

5 Results & Discussion

In our study, we employed TextBlob and the Modern Hate Speech API to analyze comments from English-language sources. Initially, we utilized language detection libraries to identify non-English comments, marking them as non-applicable for further analysis. Our results indicates that there's lot of presence of non-English comments in Youtube videos when compared to the Reddit.

For English comments, we employed TextBlob to determine the polarity score, classifying them as neutral, positive, or negative. Concurrently, we utilized the Hate Speech API to classify comments as either hateful or not. Instances of API errors during hate speech detection were logged and addressed with weekend retries to minimize data loss.

The results as plotted in Section 4.3 & 4.4 revealed a prevalence of negative tones in various sources, with a notable concentration in political subreddits. Surprisingly, hate speech was relatively minimal across the board. YouTube videos, particularly those of English singers, displayed negativity but with almost negligible hate speech overall. While the overall trend was observed, our research question delved into the specific correlation between negative sentiment and hate speech, with a focus on calculating percentage matches. Further analysis explored the relationship between sentiment and hate speech within each subreddit and each singer's videos.

Also, there's quite a difference between the cumulative of projected estimates vs actual estimates for both the Reddit & YouTube. Our initial expectation was such, YouTube is going to perform consistently and Reddit is not going to perform consistently. But, our expectation is proved to be vice versa and the actuals reveals that Reddit is consistent whereas there are only some sudden spikes in the YouTube data. Our analysis boils down to one fundamental difference we have noticed between these two platforms. YouTube users are consistent and actively engaged only when a video is just released. Unlike YouTube, the reddit users are actively engaged in a day to day usage and hence there has been a consistent trend in the overall comments.

6 Limitations & Future Scope

A significant constraint in our project lies in the treatment of non-English comments. Our current approach, relying on language detection libraries, results in the classification of a considerable portion of comments as non-applicable. This limitation poses challenges to the comprehensiveness of our analysis, particularly in understanding sentiments and detecting hate speech in diverse linguistic contexts.

To overcome the limitation associated with non-English comments, a crucial aspect of our future scope involves exploring alternative methodologies. One promising avenue is the integration of language translation libraries to convert non-English comments into English. This strategic shift not only addresses the current challenge but also presents an opportunity to expand our dataset, ensuring a more inclusive and representative analysis of sentiments and hate speech across different languages.

7 Conclusion

While we have obtained the overall sentiment tone and hate speech in the respective data sources i.e YouTube, Reddit & Politics subreddit, we are determined to find out the actual percentages of them in each subreddit and each video and show case them in our project dashboard. We will also establish how much of the percentage in comparison is hateful when the sentiment is negative.

By the end of our project we are going to answer our research questions:

1. What is the relation between the sentiment and the hate speech presence in the comments?
2. What is the overall trend in each subreddit and the YouTube video with respect to the sentiment and the hate speech presence?
3. How effective are the TextBlob and Hate Speech API against manual labelling for a certain sample size?

8 References

Noman Ashraf, Arkaitz Zubiaga, and Alexander Gelbukh - Abusive language detection in youtube comments leveraging replies as conversational context - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8507480/

https://www.ijraset.com/research-paper/analysing-sentiments-for-youtube-comments
https://medium.com/@kiddojazz/reddit-sentiment-analysis-f8a1a790124a
Understanding the Behaviors of Toxic Accounts on Reddit
https://dl.acm.org/doi/fullHtml/10.1145/3543507.3583522