Sentiment Analysis of Reddit & YouTube

Final 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 final phase of our project, we have introduced a user interactive dashboard displaying sentiment and hate speech analysis results for each subreddit on Reddit and each video on YouTube. The dashboard provides a detailed overview of toxicity levels, sentiment distribution, and the correlation between sentiment and hate speech. This approach aims to offer a user-friendly tool for understanding and navigating online communities, contributing to a more positive online environment. We have also answered if there's a relationship between the negative sentiment and the hate speech presence in the comments by undertaking hypothesis testing and eventually arriving at a conclusion.

KEYWORDS

RedditAPI, YouTubeAPI, Java, Http Client, Python, Flask framework, 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 an era dominated by digital interactions, the significance of understanding and navigating online communities cannot be overstated. With the exponential growth of platforms like Reddit and YouTube, where diverse voices converge, it becomes imperative to develop tools that foster a positive online environment. This paper introduces a pioneering approach to achieving this goal through the implementation of a user-interactive dashboard, providing comprehensive sentiment and hate speech analysis results for each subreddit on Reddit and each video on YouTube.

As a testament to the commitment to user-friendliness and accessibility, the final phase of our project integrates Flask framework for web application development. Leveraging the versatility and efficiency of Flask allows us to create a dynamic platform that not only presents analytical results but also engages users in a meaningful and interactive manner.

The visual representation of data is a key aspect of our dashboard, and for this purpose, we employed the powerful Chart.js library. Graphs and charts are essential tools in conveying complex information in a digestible format. We delve into the specifics of our utilization of Chart.js in further sections, highlighting how it contributes to the overall user experience and understanding of sentiment distribution, toxicity levels, and the correlation between sentiment and hate speech.

A pivotal aspect of our study involves addressing the relationship between negative sentiment and the presence of hate speech. Beyond a mere exploration, we have employed hypothesis testing methodologies to rigorously investigate and draw meaningful conclusions regarding this correlation. This paper unfolds the methodology employed, the data analysed, and the statistical insights gained, providing a comprehensive understanding of the dynamics that underpin online community interactions.

In summary, our endeavour revolves around the development of a cutting-edge tool that not only analyses but engages users in understanding the nuanced landscape of online communities. By shedding light on sentiment, toxicity, and hate speech patterns, we aspire to contribute to the cultivation of a positive digital space, fostering healthier and more constructive conversations in the online realm.

2. System Design Architecture for Web Dashboard

With the foundational pillars of data collection and processing seamlessly integrated into our system, the culmination of our project lies in the development of a sophisticated web dashboard for presenting our analyses. Employing Flask for backend API development ensures robust functionality and efficient data retrieval. The APIs crafted in the backend are seamlessly interfaced with the frontend, providing a unified and responsive user experience. The visualization layer is powered by Chart.js, enabling the dynamic display of sentiment and hate speech analyses through intuitive and interactive graphs. This streamlined architecture not only optimizes data flow but also ensures that our users can effortlessly engage with and comprehend the intricate insights derived from our comprehensive analysis. The accompanying architecture figure visually encapsulates the synergy between data collection, processing, and presentation in our system, showcasing the efficiency and coherence of our design.

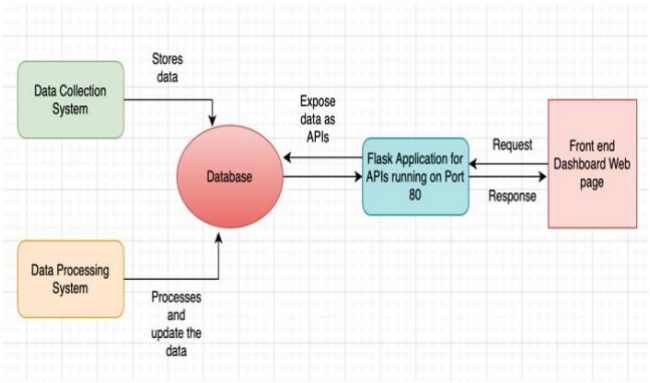

Figure 1: Web Dashboard Architecture

Web dashboard is designed to show case our results and findings. We are showcasing our 2 analyses, which will be discussed further in coming sections. Users will be able to dynamically select the date range for which the data will be visualised and is presented to them. They will be able to toggle between the data sources i.e Reddit and YouTube.

2.1 Overview of our Past Work

In the first phase of our project, we have designed a data collection system that will work like a crawler on Reddit and Youtube. We are utilising the APIs that are provided by the Reddit and Youtube to fetch the comments that are required for our project work. Below is the brief overview of how our project 1 architecture looked like.

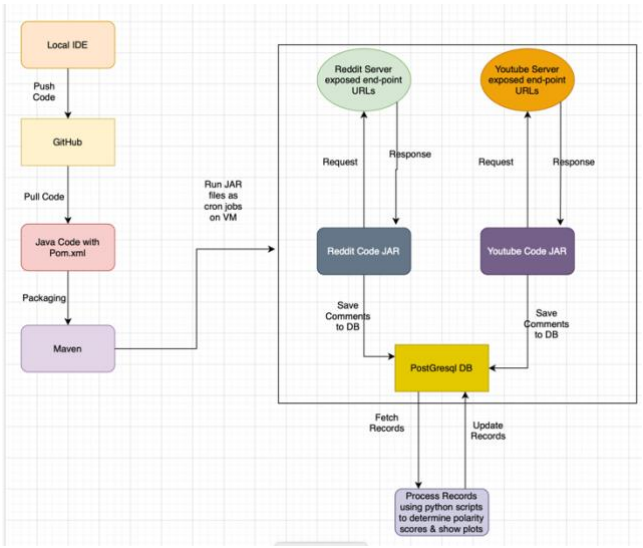

Figure 2: Data Collection System Architecture

Once the first phase of the project was setup, we went ahead with the second phase of our project i.e The Data Processing system. In this phase, we have retrieved the records that we have inserted into the system in batches and processed them to determine the Sentiment score and the hate speech value. For hate speech detection, we have made use of an third party open source API which will help us retrieve an hate score based on our input text. ModernHateSpeechAPI documentation has been cited in the references section. Following is the diagram of how our Project 2 system architecture looked like.

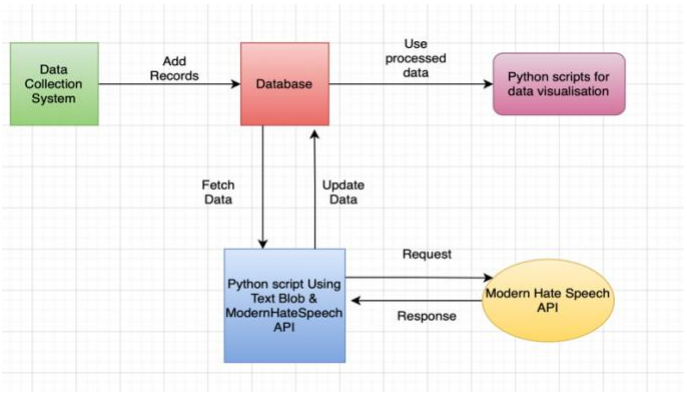

Figure 3: Data Processing System Architecture

Having arrived at the third and final phase of our project, we wanted to display our results and findings in form a user dashboard to make it easier for someone who wants to understand what's going on in the project.

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.

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

In this section, we dive deep into various ways of describing the data that we have been collecting over time. The data shown below is inclusive of both English and non-English comments & the count is subject to change as we keep collecting.

Data Source	Comments Count
Youtube	1050362
Reddit	1637673

Below is the data of the above 2 data sources in a more detailed view.

Youtube: We have collected comments from the various English singer's music video's as shown below. Singer's list includes Selena Gomez, Dua Lipa, Taylor Swift and Adele.

Video Id	Song Name	Comments count
VuNIsY6JdUw	YouBelongWithMe	346218
ixkoVwKQaJg	TakiTaki	341999
WcIcVapfqXw	CalmDown	137076
yOuqn4w1ozA	SingleSoon	59196
suAR1PYFNYA	Houdini	48972
BC19kwABFwc	LoveAgain	41675
lVkKLf4DCn8	ICanSeeYou	22611
rYEDA3JcQqw	RollingInTheDeep	22611
OiC1rgCPmUQ	DanceTheNight	18403
jDvYDzFOK9A	IDrinkWine	7800
XzOvgu3GPwY	Karma	4899

Reddit: The reddit data source contains the list of below subreddits.

Subreddit	Comments Count
r/wordlnews	267904
r/gaming	267400
r/movies	264223
r/soccer	239464
r/news	228752
r/conspiracy	205718
r/science	63736
r/gameofthrones	61125
r/breakingbad	25294
r/TrueReddit	10110
r/offbeat	3762
r/announcements	25

4.1 Data Visualisation:

The data 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 45 days.

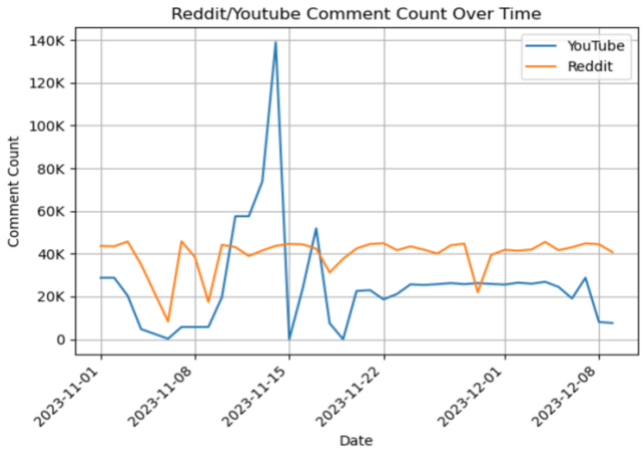

Figure 4: Total Comments on each day for last 45 days

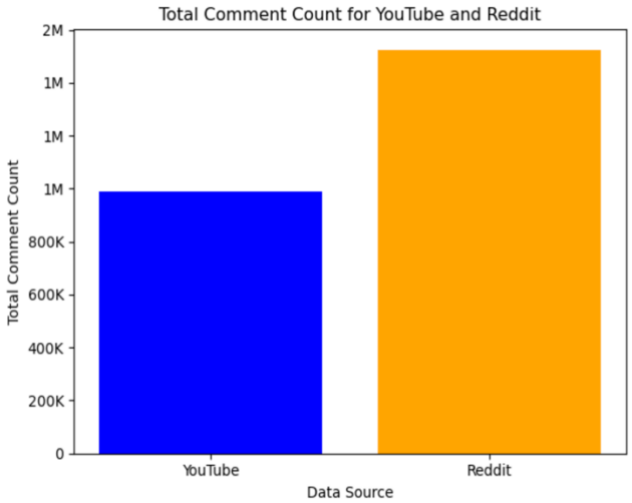

Figure 5: Total Count of comments so far

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. The graphs that are seen in the coming sections will include the English comments data as mentioned.

Data Source	Count of English Comments
Youtube	516515
Reddit	1530986

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=#
```

```
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 contain the data of how many positive, neutral, and negative comments per day in each data source since Nov-1, 2023.

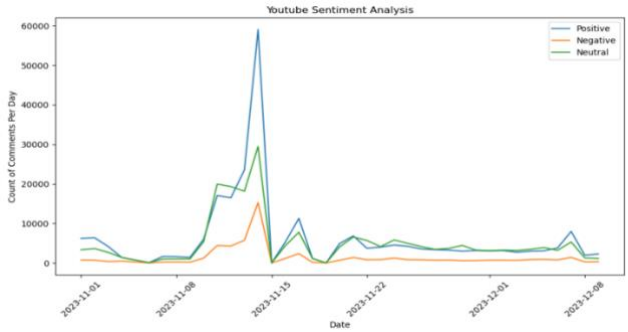

Figure 6(a): Count of Sentiment on each day for YouTube

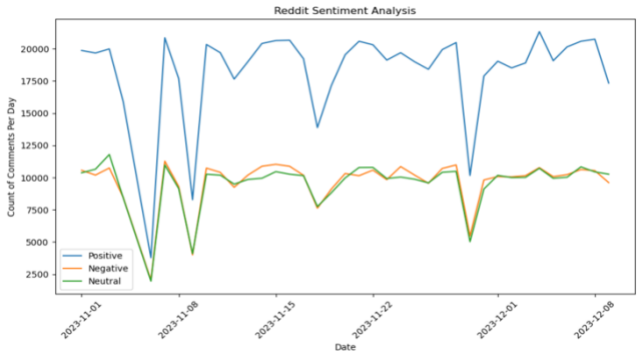

Figure 6(b): Count of Sentiment on each day for Reddit

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 two data sources. The plots below contain the data obtained from the overall list of English comments in all two data sources.

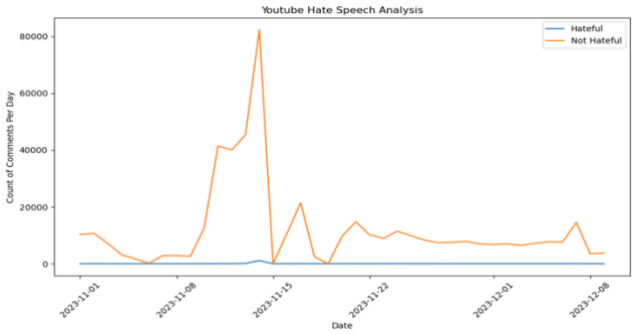

Figure 7(a): Count of Hate Speech on each day for YouTube

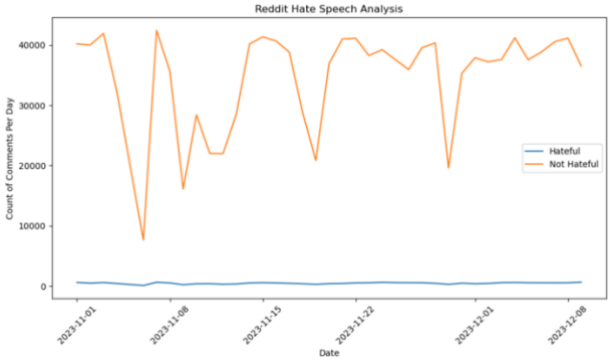

Figure 7(b): Count of Hate Speech on each day for Reddit

5.Analyses & Discussion

As mentioned at the start of the paper, in this final project we have carried out two analyses.

First Analysis: Sentiment and Hate Speech percentage in individual video Ids of YouTube Data source and subreddits of Reddit Data source.

We have tried to plot down the individual Sentiment scores of comments for each and every subreddit as shown below.

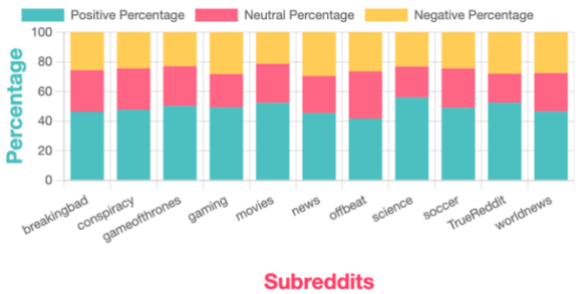

Figure 8(a): Sentiment Percentages in every subreddit

Below is the plotted data of the Hate Speech presence in each and every individual subreddit.

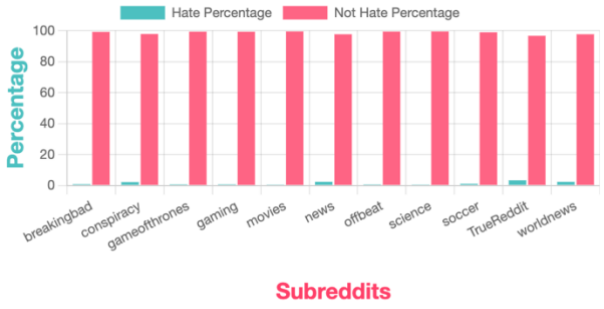

Figure 8(b): Hate Speech Percentages in every subreddit

Subsequently, below is the plot of the individual Sentiment scores of comments for each and every YouTube video that we have collected.

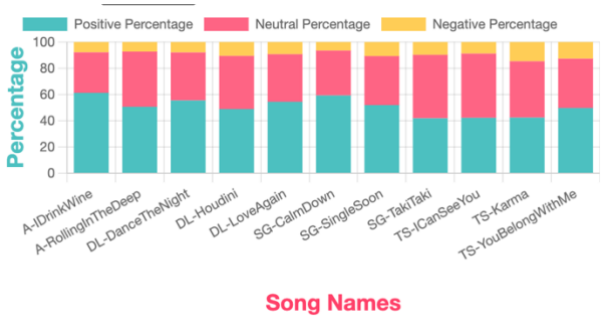

Figure 9(a): Sentiment Percentages in every Song

Similarly, here's the plot for the hate speech presence percentages in each and every video song.

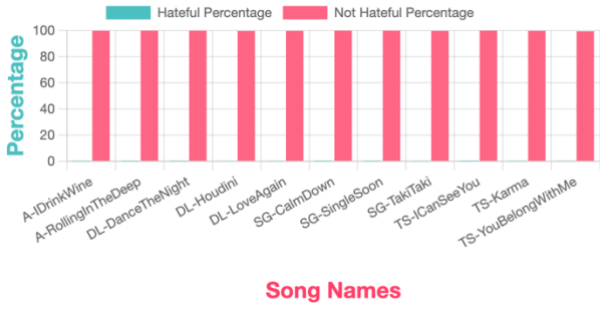

Figure 9(b): Hate Speech Percentages in every Song

A detailed comment has been provided below for the first analysis that has been performed.

Comment: As seen in figure 8a, out of all the subreddits that we have been seeing, "r/science" has more positive sentiment tone. Whereas, "r/gaming" has more negative sentiment.

Similarly, as seen in figures 9a, out of all the songs that we have been seeing, Adele's "IDrinkWine" song has more positive tone followed by the Selena Gomez's "CalmDown" song. More negative sentiment can be seen for both the Taylor Swift songs "Karma" and "YouBelongWithMe".

Overall, if you observe the figures 8b and 9b, it is evident that the negative sentiment doesn't really mean there's a hate speech presence. This will be seen in practical percentages in the second analysis.

Second Analysis: Does negative sentiment mean more hate speech ?

We have tried to find out the percentage of hateful comments from the overall "negatively" marked sentiment scores. We have felt that, this calculation can help us understand how much of the negative sentiment actually does contain hateful content. Below are our findings.

We tried to plot down percentages of Negative and hateful vs Negative and Not Hateful.

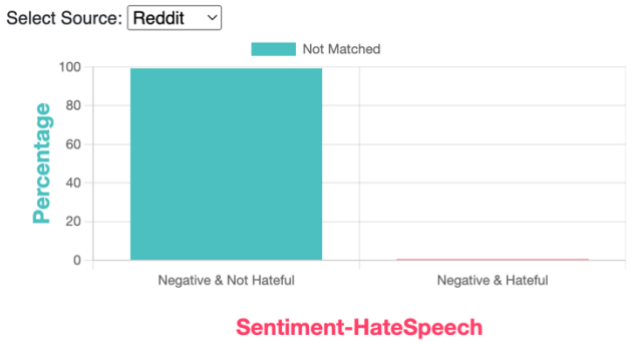

Figure 10(a): Percentages of Negative & Hateful vs Negative & Not Hateful in Reddit

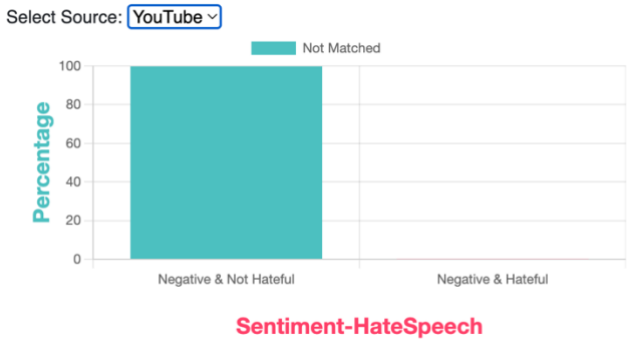

Figure 10(b): Percentages of Negative & Hateful vs Negative & Not Hateful in YouTube

Comment: As seen from the above figures, it is clearly evident that Negative sentiment has nearly 99 percent of not hateful comments. So, negative sentiment doesn't mean there is a hate speech presence.

6.Research Question

What is the relationship between Sentiment values and the hate speech presence?

We have figured out that there's no relationship between Sentiment values and hate speech presence in both the data sources i.e Reddit and YouTube. Our study has been aimed at establishing a relationship between the sentiment and hate speech presence in two different ways before arriving at our conclusion. For more details, please visit the below subsections 6.1 & 6.2.

Two methods that we have used:

1. Bar Graphs

2. Hypotheses Testing

6.1 Bar Graphs

We have clearly categorized our data into 3 categories for Sentiment: Positive, Neutral & Negative. And for hate speech: Hateful & Not Hateful. Hence, we obtained the percentages from those categories to fit into a pictorial representation of how exactly are these two related.

We have plotted down Sentiment on the X-axis as Positive, neutral and negative. Similarly, on Y-axis, we are considering the percentages of hateful and not hateful values.

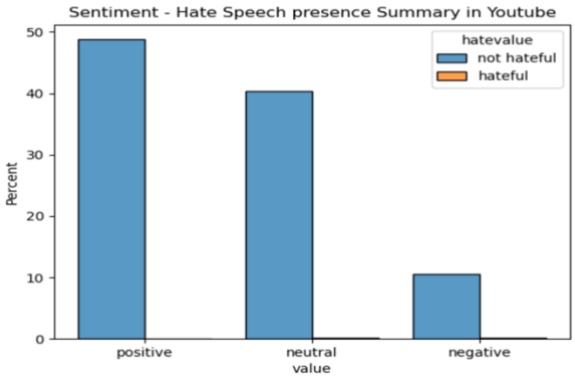

Figure 11(a): Sentiment vs Hate Speech percentage in Youtube

From the above figure, the ideal expectation is that, at least for "negative" sentiment, the hateful percentage has to be higher. But, it is clear that no significant percentage of hateful comments are present even for the negative sentiment which concludes that there's absolutely no relationship between the hate speech and sentiment values for the Youtube comments.

A similar graph has been plotted for the Reddit comments as well as shown below.

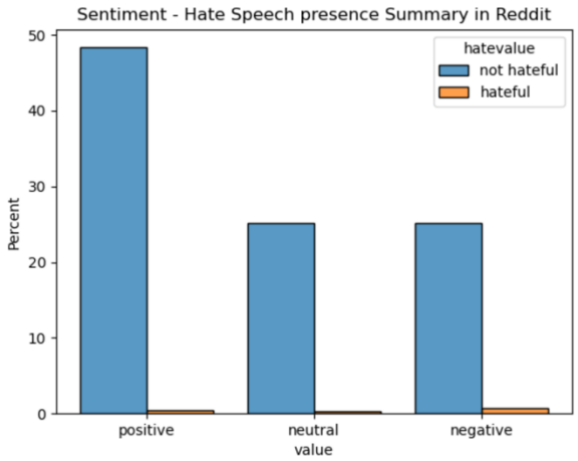

Figure 11(b): Sentiment vs Hate Speech percentage in Reddit

From the above figure, the ideal expectation is that, at least for "negative" sentiment, the hateful percentage has to be higher. But, it is clear that no significant percentage of hateful comments are present even for the negative sentiment which concludes that there's absolutely no relationship between the hate speech and sentiment values for the Reddit as well.

Comment: Although, there can be a bit of hate presence for the "negative" sentiment, it is still negligible in mathematical sense. Hence, we conclude that there's no relationship between the hate speech presence and Sentiment.

6.2 Hypotheses Testing

The Hypotheses test is a statistical method used to assess the association between categorical variables. In our study, we applied this test to explore the relationship between sentiment values and the presence of hate speech.

6.2.1 YouTube Comments:

Methodology: We conducted a test to investigate whether Negative sentiment values and the presence of hate speech are related in YouTube comments.

Hypotheses:

- Null Hypothesis (H_0): No significant association between sentiment values and hate speech labels.

- Alternative Hypothesis (H_1): Significant association between sentiment values and hate speech labels.

Results: The test produced a p-value of 1.0, which is greater than the commonly used significance level of 0.05.

Conclusion: We failed to reject the null hypothesis, indicating no significant association between sentiment values and hate speech in YouTube comments.

6.2.2 Reddit Comments:

Methodology: We replicated the analysis for Reddit comments, maintaining the same hypotheses and statistical approach.

Results: The test resulted in a p-value of 1.0, again exceeding the 0.05 threshold.

Conclusion: Consistently, in both YouTube and Reddit datasets, we found no evidence to reject the null hypothesis. This suggests no significant association between sentiment values and the occurrence of hate speech.

Our robust statistical approach strengthens the validity of our conclusion, providing compelling evidence against the independence of sentiment values and hate speech presence across both data sources.

7.Limitations & Future Scope

In hindsight, we have realised that there's a unintended limitation in our project. That is, we are not storing the toxic values of ModernHateSpeech API in the database. While trying to process the data in the data processing stage, we are making use of those toxic scores to categorize them and those categories like "hateful" and "not hateful" are being stored in the respective columns of the database. This has been a limitation for us when we wanted to put the corelation graphs between Sentiment and Hate Speech presence. Had we got the numerical values, we would have been able to try fitting a curve using those numerical values. However, in future, we would like to run the records in batches and store the numerical values as well.

Another 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.

One final but a simple constraint would be the limitations of chart js library. We were able to map the values on X-axis and Y-axis but not able to put a common label name for each X-axis & Y-axis. We are yet to full explore the features of the library and we are determined to fix it.

8.Conclusion

Having obtained the overall sentiment tone and hate speech in the respective data sources i.e YouTube & Reddit, we have aimed at contributing to a more positive online community interactions and we believe we are successful with our attempt. Besides trying to label the hateful content in the respective online communities, we have also conducted our studies to establish a relationship between the sentiment values and the hate speech presence and eventually we have figured out that there's no relationship between the two. Our future studies will focus at employing more robust methods to clean and process the data to be more accurate at determining the hate speech presence on social media platforms.

9.References

[1] 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/

[2] "Analysing Sentiments for YouTube Comments." https://www.ijraset.com/research-paper/analysing-sentiments-for-youtube-comments

[3] Kiddojazz. "Reddit Sentiment Analysis." Available at: https://medium.com/@kiddojazz/reddit-sentiment-analysis-f8a1a790124a

[4] "Understanding the Behaviors of Toxic Accounts on Reddit."- https://dl.acm.org/doi/fullHtml/10.1145/3543507.3583522

[5] Mollas I, Chrysopoulou Z, Karlos S, Tsoumakas G. "Ethos: an online hate speech detection dataset." ArXiv preprint, 2020. -https://arxiv.org/abs/2006.08328