Sentiment Analysis of Reddit & YouTube

A User-Interactive Dashboard Proposal

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 aim to introduce 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.

KEYWORDS

Data Visualisation, Python, Flask Web Framework, Chart.js, HTML, Javascript.

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 initial phase of our project, we successfully established a robust data collection system that continuously gathers data from the Reddit and YouTube APIs. This comprehensive dataset forms the backbone of our investigation into addressing negativity and toxic behaviour on these prominent social media platforms.

Building upon this foundation, the second phase of our project involved extensive data processing. We applied sophisticated techniques to extract sentiment and hate speech scores from the comments continuously streaming into our collection system. This crucial step allowed us to quantify the emotional tones and identify instances of toxicity within the online discourse on both Reddit and YouTube.

As we enter the third and final phase of our project, our focus shifts towards delivering a user-friendly and interactive solution. Our objective is to present the accumulated insights through a comprehensive dashboard. This dynamic tool will address two fundamental analyses, providing users with an intuitive interface to explore and comprehend the intricate dynamics of sentiment and hate speech within various online communities.

2 Methodology & Implementation plan

This section will outline about the research question that we are planning to answer, the two analyses that we are going to perform for our interactive dashboard and the tools, libraries that we are planning to use for the project implementation.

2.1 Research Question

The one research question that we are going to answer would be, "Is there a relationship between Sentiment scores and presence of hate speech in the Reddit & YouTube comments?". We are going to provide a detailed explanation and analysis from our findings.

2.2 Interactive Tool Analyses

The interactive tool facilitates a user-friendly experience with the ability to select a date range directly from the dashboard, presenting dynamic line plots illustrating comments collected during the chosen period.

Line Plots Based on Date Range: Users will be dynamically presented with the comment's distribution over the time between the selected date range.

First Analysis: Sentiment and Hate Speech Percentage:
Explore sentiment and hate speech percentages for each subreddit and YouTube video in the chosen date range, offering insights into emotional tones and toxicity levels across different online communities. User will be presented with the respective plots for the same.
Second Analysis: Relationship Between Sentiment and Hate Speech:
Examine the percentage correlation between sentiment scores and the presence of hate speech within each subreddit and YouTube video during the selected date range. Gain a deeper understanding of how negative sentiments relate to the prevalence of hate speech in online spaces. User will be presented with the respective percentage plot for the same.

2.3 Tools for Implementation

In our implementation, we intend to leverage the Python Flask web framework to construct the APIs that power our application. For the visual representation of data through plots, our plan is to employ the chart.js library.
It's important to note that the selection of libraries is tentative and subject to change during the actual implementation. Any deviations from the current plan will be reflected in the final report, which will provide an accurate account of the tools utilized in the project.

3 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/abs/10.1145/3543507.35835