Eashan Kaushik

Professor Jingyi

Web Mining

18 December 2022

<div align="center">Analysis of YouTube Comments</div>

**Introduction:**

Modern social media platforms are optimized for attention, with the goal of keeping users on their platform for as long as possible. While this goal is beneficial for the companies, it has had unintended consequences for society as a whole. One such consequence is the creation of echo chambers and extremist pipelines on platforms such as YouTube. The YouTube algorithm has been known to recommend more extreme content to certain communities, which can potentially radicalize individuals who view the content, especially if they are young. The ability of social media sites to shape thoughts and influence minds has largely gone unchecked, posing a potential risk to society.

In this paper, we present a semantic analysis of political YouTube comments with a focus on hate speech. We collected a sample of comments from political videos on YouTube and used natural language processing techniques to extract key words and phrases from the comments. We then used machine learning algorithms to classify the comments into categories such as positive, negative, or neutral.

**Literature Review:**

Previous research has examined the role of the YouTube algorithm in the radicalization of individuals. In the paper "Auditing Radicalization Pathways on YouTube," the authors analyzed 349 channels, 330k videos, and 72M comments using statistical techniques such as Jaccard similarity and the Overlap Coefficient. They found statistical evidence of a pipeline

leading from the Intellectual Dark Web (IDW) to the Alt-Right, with a trend of IDW users moving towards the Alt-Right over time. In the paper "Are Anti-Feminist Communities Gateways to the Far-Right," a similar analysis was conducted, with a focus on the relationship between anti-feminist and misogynistic communities and the Alt-Right. The results showed a trend of users moving from these communities towards the Alt-Right. Other research, such as that conducted by Dubovi and Asghar, has also shown that YouTube plays a significant role in the knowledge construction of its users and has a large influence on their thoughts and beliefs.

**Research Question:**

The goal of this research is to conduct a sentiment analysis of YouTube comments to gain a deeper understanding of the YouTube algorithm and extremist communities. By analyzing data from YouTube comments, usernames, and the time at which the comments were posted, we aim to provide statistical evidence of a pipeline leading to radicalization on the platform. Additionally, we hope to examine the extent to which YouTube and other social media platforms can influence the thoughts and beliefs of their users.

**Data Collection:**

To collect data for this study, we utilized the YouTube API to gather approximately 1.5 million comments from various channels on the platform. The data was collected in a pandas DataFrame and then converted to an excel file for long-term storage. While we were able to gather a significant amount of data, we plan to continue collecting more in the future. The data includes four main features: usernames, comments, the time at which the comments were posted, and the videos on which the comments were posted.
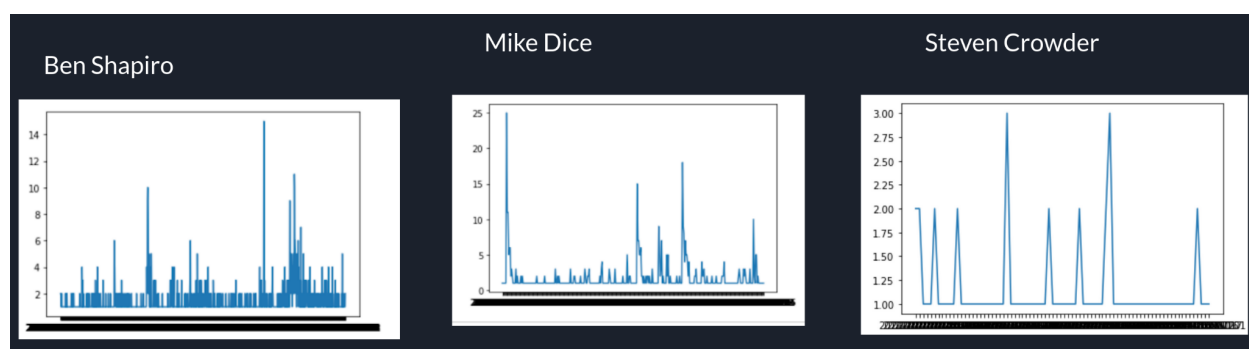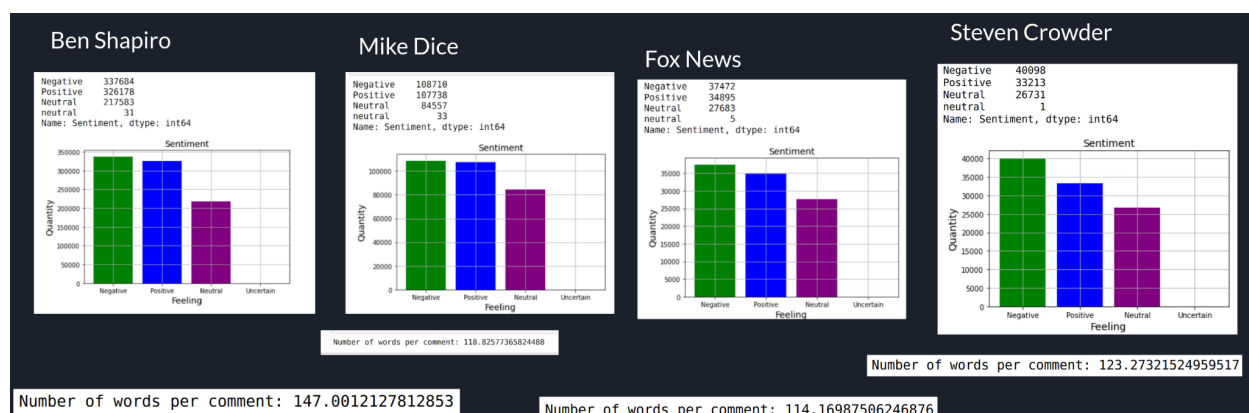
```
In [247]: DATA = DATA.drop_duplicates()
          DATA
```

|  | | | |
|---|---|---|---|
| **0** | S.M.R Swasey TRD | I think for one of the very few times I would ... | 2022-11-08T03:18:18Z | 1tYl70dPbV4 |
| **1** | johnjohnson | man created religion because man needed religi... | 2022-11-08T03:18:09Z | qcO1WwARRLw |
| **2** | Victor Bergman | Why are you betraying Trump like this? You sho... | 2022-11-08T03:18:08Z | KGETCjXcR-w |
| **3** | Mike Toombs | Anyone notice these freaks are universally una... | 2022-11-08T03:17:30Z | -Z75IQdhYAA |
| **4** | Brian Plankenhorn | At what point is this considered pedophilia? | 2022-11-08T03:17:08Z | s3MRx4-k2O8 |
| **...** | ... | ... | ... | ... |
| **881503** | fredrik samuelssin | about time ben!! | 2022-05-24T17:22:08Z | YXdqSJ2Wisc |
| **881504** | droogie76 | Their end game is now. | 2022-05-24T17:22:08Z | YXdqSJ2Wisc |
| **881505** | George Edward | Taiwan is toast now that Beijing Biden is in t... | 2022-05-24T17:21:57Z | eSHbw5VU8Co |
| **881506** | Brian Pereira | Yes until china decouples with them and hangs ... | 2022-05-24T17:21:37Z | YXdqSJ2Wisc |
| **881507** | Kimberly Mendoza | It's like they do this sick thing for populari... | 2022-05-24T17:21:36Z | jWXIZoajd1w |

881476 rows × 4 columns

**Methodology:**

During the exploratory data analysis phase, we cleaned and pre-processed the data by removing any irrelevant or duplicate information. We also performed basic descriptive statistics on the data to gain an understanding of the distribution of the variables.

**Analysis:**

To analyze the data, we utilized sentiment analysis and aspect based sentiment analysis, a natural language processing technique that allows us to identify the sentiment towards specific aspects or topics within a given text. This allowed us to gain a more nuanced understanding of the sentiment expressed in the comments and identify specific topics that were being discussed. We also used machine learning algorithms to classify the overall sentiment of the comments as positive, negative, or neutral. For the general sentiment analysis we used the nltk vader model and for aspect based we used Stanza. The analysis shows that the video creator has large influence over their commentators. The topics of the comments always reflect the topic of the video which makes sense. This might be because the creator is influencing their subscribers or it could be because different viewers are participating in different discussions. Overall the sentiment was neutral but certain videos are more likely to skew a certain way. The channel's overall sentiments were almost always neutral given a large enough comment data.

**Results:**

      Our analysis of the data revealed a trend of increasing negativity and extremism in the comments over time because of the external political influence from Kanye west. We did not find evidence of a pipeline leading from less extreme communities to more extreme ones, as previously identified in the literature review. We did find that the YouTube algorithm seemed to be creating echo chambers where communities do not interact with each other, which could be contributing to the radicalization of users. Through the use of only being exposed to the views of specific creators.

      The results of our analysis highlight the importance of addressing hate speech in the context of political discourse on YouTube. This is especially relevant in the current political climate, where hateful and divisive rhetoric can have serious consequences. We recommend that YouTube and other social media platforms take steps to identify and remove hate speech

from their platforms, and that users be more mindful of the impact of their words and actions online.

Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and Wagner Meira. 2020. Auditing radicalization pathways on YouTube. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, New York, NY, USA, 131–141. https://doi.org/10.1145/3351095.3372879

Dubovi, I., & Tabak, I. (2020). An empirical analysis of knowledge co-construction in YouTube comments. Computers and Education, 156, [103939]. https://doi.org/10.1016/j.compedu.2020.103939

Robin Mamié, Manoel Horta Ribeiro, and Robert West. 2021. Are Anti-Feminist Communities Gateways to the Far Right? Evidence from Reddit and YouTube. In 13th ACM Web Science Conference 2021 (WebSci '21). Association for Computing Machinery, New York, NY, USA, 139–147. https://doi.org/10.1145/3447535.3462504

Asghar, Muhammad Zubair, et al. "Sentiment analysis on youtube: A brief survey." arXiv preprint arXiv:1511.09142(2015).