Sentiment Analysis on Youtube Comments

Author: Zixuan Ge, zge7@

Name: Zixuan Ge, zge7
Captain: Zixuan Ge
Group Name: dontYellAtMe
Github Repository: https://github.com/marcuzge/CS410CourseProject

Have you always wondered what other people thought when your favorite YouTuber posted a video on YouTube? Have you wondered how the public reacted after seeing the news videos that Liz Truss pushed tax cuts for the rich? Have you wondered when the Super Bowl Halftime show's performer is announced on YouTube, are people sending out more positive or negative comments? I plan to solve these questions with a command-line tool that does sentiment analysis on YouTube comments for a specific video. Users can run a command in the terminal, providing the url of the YouTube video as a parameter and get back directly whether this YouTube video's comments are leaning positive, negative or neutral. A pie graph will also be generated at runtime to provide better visualization. This is interesting as I always wondered when a controversial video comes out, how others would react on YouTube. This is important because YouTube removed the public dislike count back in November 2021. While content creators can still view this in the background, viewers can now only see how many likes a video has received. I will use YouTube data API, NLTK's Vader sentiment analysis tools, Pandas for data cleaning, analysis and pie chart generation. My planned approach is roughly the following:

1. Get familiar with YouTube's data API to scrape comments from YouTube videos (https://developers.google.com/youtube/v3/getting-started)
2. Use Panda to create dataframes for YouTube comments and perform data cleaning (https://pandas.pydata.org/)
3. Use NLTK's Vader sentiment analysis toolkit to perform sentiment analysis (https://www.nltk.org/_modules/nltk/sentiment/vader.html)
4. Display the text result back to the user in the Terminal
5. Use Panda to create Pie Chart for better visualization

I will use mostly Python to finish this project. The expected outcome includes both text results and a Pie Chart displaying the percentages of positive, negative and neutral comments for a

YouTube video. It will be evaluated by the accuracy of the sentiment analysis on the YouTube video's comments. I will collect data from human evaluators and compare it with my result.

The main tasks to be completed are listed above. As I am new to the YouTube data API, I expect to spend 10 hours reading documentation, setting up a test environment and scraping comments from the video. Then, I will spend 3 hours cleaning and creating "comments" dataframes with Panda. To become familiar with Vader and perform sentiment analysis with the toolkit, I expect 8 hours of work. To display the data back to the user I will use 3 hours to complete. This does not include debugging time and is already above the 20+ hours requirement of work for 1 person.