



Determining how popular your favorite artist is on Reddit

Julia Bighetto, Kyle Bryant
CS591 L1



Problem:

In a world where people can voice their opinions in various channels in the internet, it is easy to lose track on specific information you're looking for. If you are trying to keep track of someone's reputation on a website, for example, you would have to scour and analyze every comment in regards to that person to determine if it is a positive or negative comment.

Relevance:

This problem is relevant and interesting because we are dealing with data scraping, analysis, manipulation. Therefore, our problem is relevant because it delivers a solution to the data-analysis problem in a major social media platform. This could also be further implemented for other social media platforms

Furthermore, this could lead to a more complex project. While here we will analyze only keywords, we could improve our algorithm by building an environment that understands a human's intent (without specific keywords).

Objective:

The objective of this project is to analyze an artist's reputation/popularity in Reddit (in general or specific subreddits) by analyzing comments that are in direct relation with said artist and determining the popularity via keyword analysis.

Key word: amazing
Weight: 0.9

Key word: offensive
Weight: -0.6

Travis Scott is an **amazing** artist. Some people think his songs are **offensive**, but I don't care. He's **brilliant**.

Key word: brilliant
Weight: 0.9

Total weight: $0.9 + -0.6 + 0.9 = 1.2$

Methods:

Setting up our project:

- Decided on 20 different genres
- Scraping Wikipedia for all artists per genre via the requests API and mwparserfromhell API
- Cleaned the data to avoid gibberish input via mwparserfromhell and regents API
- Found a list of positive and negative words and assigned a value between -1 and 1 (-1 being a strong negative word and 1 being a strong positive word)

Parsing through reddit:

- Using relevant subreddits to each genre, we pull all comments from posts relating to (directly or indirectly) to that artist.
- From there we are able to create a list of strings and eliminate any useless words from each.
- We can then input it into our algorithm.

Creating a simple User Interface

- We used Tkinter for our user interface due to its simplicity, and we wrapped it with PySimpleGUI to make it look better.
- The interface has four options and the ability for you to see our entire artist list.

The algorithm:

- Using the Reddit API to gather comments about the artist/genre the user selected, we use the positive/negative words dictionary to go through the comments and assign a weight on each comment.
- Once we are done parsing through the comments, we sum up the weights and see the artist's popularity.

Key word: awful
Weight: -1

Key word: boring
Weight: -0.5

I would never hear any 21 Pilot's songs if I could. They're **awful**, **boring**, and I just can't stand the guy's voice.

Total weight: $-1 + -0.5 = -1.5$

By the numbers:

8130 artists

330 positive/negative words

20 genres

Results:

