Xuan Huang
Mar 31, 2020

SI 507 — Final Project Proposal

Overview

My proposed project is a program that evaluates specific characteristics such as sentiment and toxicity of text comments made by users on social media sites.
It will allow a user to look up a username on a specific social media platform (i.e. Twitter and Reddit), and utilizes a text analysis AI to display sentiment, emotion, and abuse characteristics for the last 50 comments posted under that username.

Data Sources

- ParallelDots Text Analysis APIs (https://www.paralleldots.com/text-analysis-apis)
- Twitter API (https://developer.twitter.com/en)
- Reddit API (https://www.reddit.com/dev/api)

Challenge Scores

Source	Source Type	Auth Protocol	Score
ParallelDots API	Web API you haven't used before that requires API key or HTTP Basic authorization	API key	4
Twitter API	Web API you've used before	OAuth 1.0a	2
Reddit API	Web API you haven't used before that requires OAuth	OAuth 2.0	6
		Total Challenge Score	**12**

Access Attempts

Developer accounts were created on each API site, and API keys were obtained from all sources. For Twitter and ParallelDots, there was minimal difficulty in doing so. For Reddit, the process for extracting content varies by the type of content. While I was not yet successful in parsing user comments, I did find multiple resources for doing so by importing **`PRAW`** (https://praw.readthedocs.io/en/latest/tutorials/comments.html).

Data Presentation

As the ParallelDots has the capability to output quantitative results from text inputs, the data obtained from this program will be presented through various tables and graphs with Plotly. For example, once a user has inputted a username and social media platform (either Reddit or Twitter), they will be presented with various graphs displaying the average sentiment, emotion, and abuse values for all comments obtained via the API, as well as specific comments that scored extremely high or low on a specific metric (for example, their comment with the highest "anger" score).