**Programming: Text Classification**

**Tools Used**

Web Scraping: Beautiful Soup

Named entity recognition: spaCy, NLTK

Sentiment analysis: TextBlob and VADER

Topic modeling: sklearn and Gensim

**Results**

A bar graph with blue bars

Description automatically generated

**Analysis/Reflection**

I scraped approximately 200 health related articles from the time.com news portal. From named entity recognition, the top entities were organization, date, and person. This makes sense since these entities are frequently used in news articles. Some most common entity labels were: covid-19, FDA, CDC, WHO, Pfizer, Moderna, etc. They are frequently mentioned as many articles discuss various aspects of the ongoing pandemic, from vaccine development to public health measures and global efforts to control the spread of the virus.

The sentiment analysis resulted in overall more positive sentiments (134) than negative sentiments (65). I was expecting there to be more negative sentiments than positive sentiments due to the presence of the pandemic. However, after performing topic modeling, it seems like a lot of the articles centered around the topic vaccine