

---

# Webscraping Reddit

## Sentiment Analysis

Alex Lue • CMPS 184

---

# sentiment analysis

## opinion mining

- a part of nlp that is a systematic way to process subjective information
- used for movie reviews, political views, marketing, etc

---

---

# goals

## webscrape subreddits

- using PRAW to collect post data from subreddit of choice

## perform SA on our data

- specifically it will be on the posts, including their text body, process it in a way that can be used by VADER
  - use that to build our own ML model for sentiment analysis
-

---

# process

## reddit

- choose the subreddit, use praws to collect specifically what info you want from that page into a dictionary
- export to a pandas dataframe

## sentiment

- for each entry in the dataframe from praws, use vader to collect the sentiment and add that back into the dataframe
  - use a nltk to help format sentences to remove common stop words to make a word frequency chart
-

---

# libraries used

## reddit

- PRAWS
- pandas/matplotlib

## sentiment analysis

- VADER
  - nltk
  - pandas/matplotlib
  - scikit learn
-

---

# links

- <https://towardsdatascience.com/scraping-reddit-with-praw-76efc1d1e1d9>
  - <https://praw.readthedocs.io/en/latest/index.html>
  - <https://github.com/cjhutto/vaderSentiment>
  - <https://medium.com/@vasista/sentiment-analysis-using-svm-338d418e3ff1>
-