

Sentiment Analysis of r/dogecoin, r/CryptoCurrency and stock price

By: Matthew Padgett, Danielle White, Ciara Spencer, and Alex Goldstein

Project Mission: Run Sentiment analysis on subReddit comments from cryptocurrency and compare sentiment with stock price over a 3-month period.

Question - Does overall subReddit sentiment correlate with change in the stock price ? (e.g. positive/negative sentiment vs high/low)

Project Data Sources:

<https://www.reddit.com/r/dogecoin/>
<https://www.coingecko.com/en/api>
<https://github.com/cjhutto/vaderSentiment>
<https://github.com/rhiever/reddit-analysis>
<https://www.specrom.com/reddit-wordcloud-generator/>

Project Methods:

Run sentiment analysis using VADER tool on subreddit cryptocurrency using PRAW (reddit API)
Run data using train / test in Naive Bayes model to compare results to VADER model
Run comparison analysis against day to day stock price over the course of 3 months using coingecko API based on hour to hour interpreted sentiment.

Alt-Options / Add-ons:

Option - comparative analysis of sentiment values based on two subreddits in regard to cryptocurrency cross-analyzed with stock price within date range, i.e. r/dogecoin vs r/cryptocurrency (further analysis of data, comparison)

Option - wordCloud of subReddit to modify VADER lexicon sentiment specific to sub vernacular (improves sentiment)

Requirements:

- Machine learning library
- At least 2 of the previously-learned tools
- Host application
- 15 minute presentation

AWS for hosting of subReddit comment data

VADER

PRAW (reddit API)

CoinGecko (crypto API)

Heroku Deployment

Challenges:

Figuring out how to account for a lag in price changes and reddit posts

Converting timestamps into month/day/year format

Scraping reddit data

Time it takes to run a Jupyter Notebook that parses through thousands of reddit comments

Accounting for missing threads

For presentation:

Title slide w/ names

Project mission and question we hope to answer

Project methods/tools used

Visuals (line chart)

Conclusions

Wordcloud

Challenges

List sources