

Tech Review

Can monitoring r/wallstreetbets subreddit predict a meme stock surge?

I. Abstract

End of year 2020 and beginning of 2021, r/wallstreetbets subreddit members were orchestrated by a group of retail stock holders caused phenomenal stock rallies for two tickers GME, AMC. There was no real performance in those two companies which led to their stock surges. The phenomenon simply occurred due to great orchestration via meme in social media. There never been one or many true factors that can absolutely predict stock price trend, this tech review is purely an attempt to do so by looking into historical data of subreddit comment and stock price at the time to find the correlation. Based on that correlation, a prediction may be achieved.

II. What available tool can make the prediction?

1. Public stock data from Yahoo Finance API.
 - a. What data the API provides for a ticker
 - Date
 - Open (price)
 - Close
 - High
 - Low
 - Volume

There are a few more parties which provides free/subscription-based services where we can pull stock historical data and/or live data. For example,
[Yfinance API - Realtime & Historical Data API - polygon.io](https://polygon.io/yfinance)

- b. Sample retrieval data

```
{
  "AMC": {
    "2020-01-02": {
      "stock_data": {
        "Open": "7.3",
        "High": "7.56",
        "Low": "7.23",
        "Close": "7.46",
        "Adj Close": "7.414512",
        "Volume": "4545900",
        "Change": "0.160000000000000014",
        "Change Rate": "0.021917808219178103"
      },
      "reddit_data": {
        "submissions": []
      }
    },
    "2020-01-03": {
      "stock_data": {
        "Open": "7.45",
        "High": "7.6",
        "Low": "7.31",
        "Close": "7.32",
        "Adj Close": "7.275366",
        "Volume": "2218000",
        "Change": "-0.12999999999999999",
        "Change Rate": "-0.017449664429530186"
      },
      "reddit_data": {
        "submissions": []
      }
    },
    "2020-01-06": {
      "stock_data": {
        "Open": "7.25",
        "High": "7.3",
        "Low": "7.11",
        "Close": "7.12",
        "Adj Close": "7.076585",
        "Volume": "2903400",
        "Change": "-0.12999999999999999",
        "Change Rate": "-0.017931034482758606"
      },
      "reddit_data": {}
    }
  }
}
```

2. Reddit API

a. Based on an input subreddit, it can provide

- Author
- Score
- Upvote_ratio
- num_comments
- comments (list)
 - o author
 - o body
 - o score

b. Sample retrieval data

```
documents_2tickers.json
{
  "AMC": {
    "01Jan2020": {
      "reddit_data": {
        "submissions": [
          {
            "id": "1577922814.0",
            "self_text": "[deleted]",
            "score": 7,
            "upvote_ratio": 0,
            "num_comments": 0,
            "comments": [
              "I don't know bro, I think you are gay",
              "I called Blue when it was at 72. I don't think these two presentations that you're mentioning are gonna involve any data, so idk if they'll provide a price boost",
              "If there's something I learned on this sub, it's that you should always reverse the advice. \n\nTherefore, YOLO short $BLUE.",
              "Holy shit, I thought this sub was only for retards but my friend, you are clearly an autistic.\n\nGoing all in tomorrow",
              "This is the only publicly traded company I've ever seen with all lowercase letters",
              "Trying to sell your"
            ]
          }
        ]
      }
    }
  }
}
```

III. Implementation

1. Obtain historical data
2. Pull historical 'meme' stock data.
3. Pull historical reddit comment about 'meme' stock
4. Study correlation between stock price and high score, related comment posted in Sub-Reddit 'wallstreetbet' based on date chronologically

IV. What left to be done to improve the prediction?

Since learn correlation between stock price and r/wallstreetbet subreddit by using its historical data may form a good prediction. Spotting a surge in meme stock ahead of time

can be very tricky. It may require more data sources and activities to improve quality of the prediction. Some ideas as below.

- Pull data from Google Trends
- Pull data from News
- Watch *r/wallstreetbet* subreddit in real-time

V. Sources

- Yahoo Finance API

<https://reasonabledeviations.com/2017/07/30/yahoo-historical-prices/>

- Reddit API

<https://praw.readthedocs.io/en/stable/>

<https://towardsdatascience.com/how-to-use-the-reddit-api-in-python-5e05ddfd1e5c>

- Screenshots of retrieved data is from *credited to* team yoda

VI. References

- Yahoo Finance API - A Complete Guide - AlgoTrading101 Blog

<https://algotrading101.com/learn/yahoo-finance-api-guide/>

- Google Trends

Unofficial API for Google Trends <https://pypi.org/project/pytrends/>

- Newscatcher

<https://github.com/kotartemiy/newscatcher>