FinTwitter

A tool to analyze real-time company stock buzz on twitter and identify future “black swan” events similar to GameStop

March 11, 2021

Prepared by: Noah Zuckerman

Problem and hypothesis

Over the course of our DSI class, a new phenomenon was introduced into the world of finance. One name for it would be **distributed crowd consensus stock trading**.

Add a little bit of virality to the equation and GameStop became the first of a very new breed of stock market trade. This new movement is unique within the history of the stock market as it has redditers, tweeters and average retail traders sitting in the drivers seat for the first time ever. In response to this, new ETFs that track stocks based on their social media sentiment have launched.

The investment hypothesis is simple: people are transparent and discussing their trades through social media. Aggregate all this data up to the highest level and leverage the “wisdom of the crowd”.

The purpose of my capstone project is to better understand the network effects of this phenomenon, test the hypothesis and build a prototype “early warning system”. The key deliverable from the project is tool that scrapes tens of thousands of tweets daily to analyze and model the relationships between **frequency of mentions, level of engagement and stock prices** for publicly traded companies.


Search

Bloomberg Opinion

Money Stuff

The GameStop Game Never Stops


CNBC MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV WATCHLIST PRO

PERSONAL FINANCE

More bubbles, less shorting. What the GameStop craziness could mean for the future of investing

PUBLISHED SAT, FEB 6 2021•9:45 AM EST | UPDATED WED, FEB 10 2021•9:49 AM EST

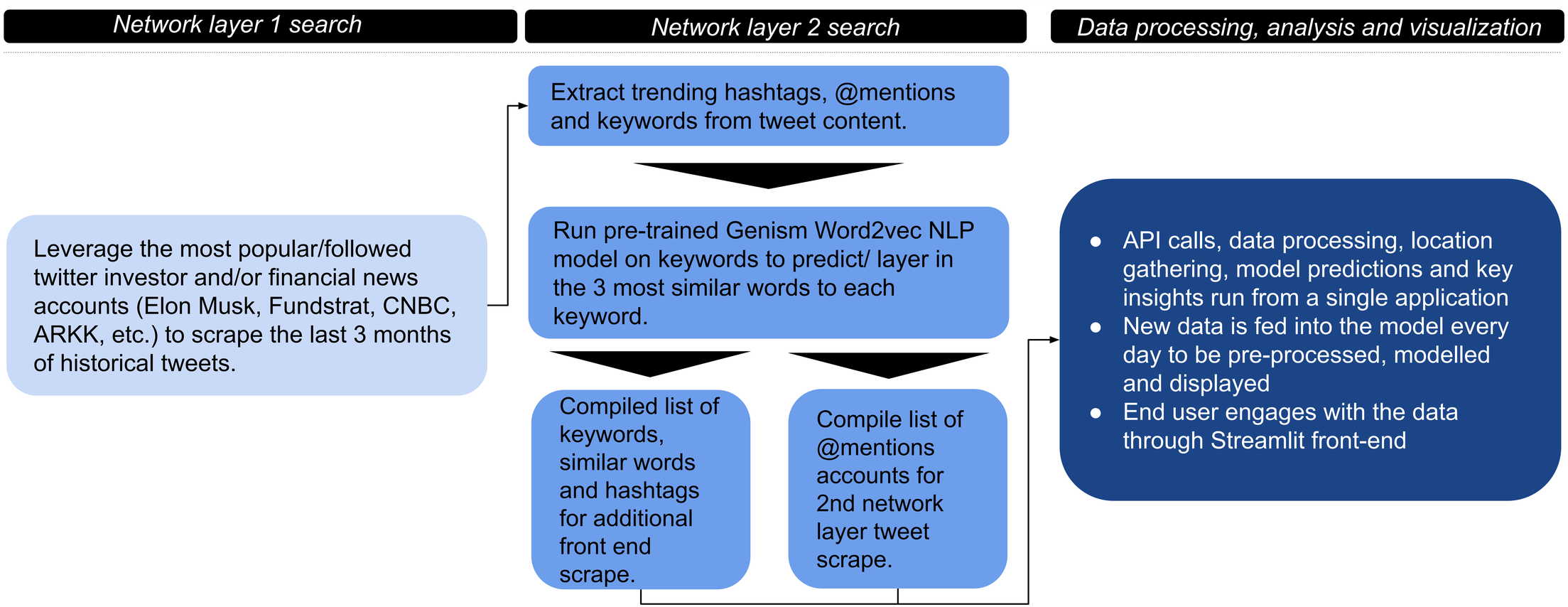

CNBC MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV WATCHLIST PRO

TRADER TALK

‘Buzz’ ETF tracking social media sentiment launches Thursday amid Reddit manias in stocks

PUBLISHED THU, MAR 4 2021•7:35 AM EST | UPDATED THU, MAR 4 2021•10:33 AM EST

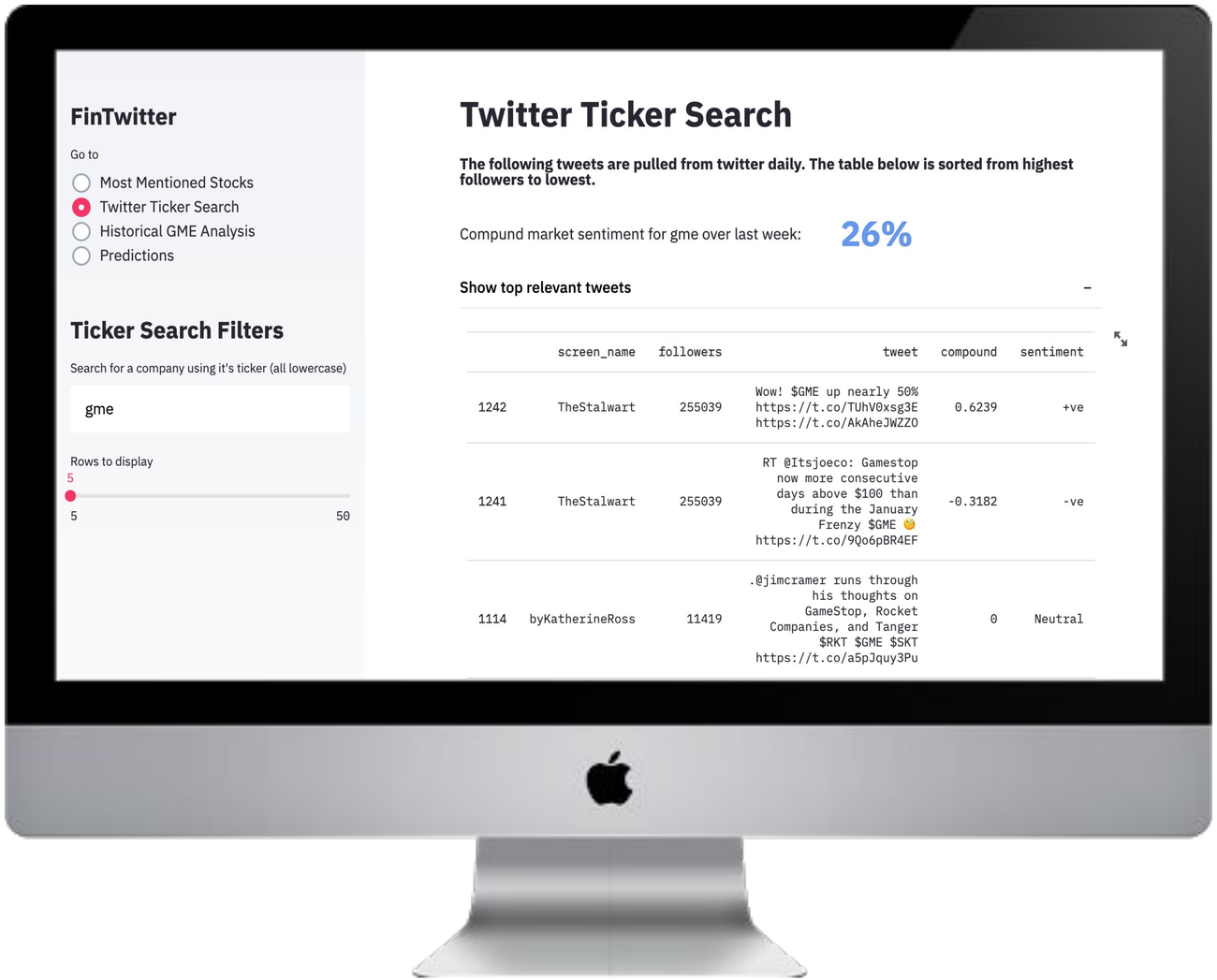
How does it work?



Tech Stack

Python 3.8 | Genism | Tweepy | Jupyter | Streamlit | Heroku

FinTwitter Demo!



Limitations, risks and future improvements

Limitations and risks	Mitigations and/or improvements
Twitter data processing only counts a company mention if it matches the stock ticker (not the name of the company itself)	Build-up a repo/database of common company names / spellings and leverage this in mention identification logic
Time, processing-power and storage limitations only allowed me to pull 500K historical tweets. Relatively small sample compared to the data that is available out there.	Build a proper back-end storage system and migrate processing modules to more powerful AWS CPU instance to optimize for speed.
Number of previous “black-swan” events to study and model is limited to only GameStop data.	Although model predictions are interesting and seemingly correct this week. Given the limited historical data, I would take them with a grain of salt and only use them as an input to a fully formed qualitative investment decision until the model is trained on more data.