



Tiago Nascimento

USING REDDIT FOR TRADING

SCRAPING & SENTIMENT ANALYSIS

Investment Strategies



Passive: Index funds

- historically outperforms +80% of active managed funds
- low fees
- lower risk
- low effort required

Active: Stock picking

- bigger return potential
- low fees
- higher risk
- significant effort required



STOCK PICKING APPROACHES

Fundamental

Seeks value opportunities by assessing the overall economy, industry conditions, companies' financial strength, competitive advantages, management, perspective, etc.

key metrics: P/E Ratio, dividend yield, operating cashflow

Technical

Seeks opportunities by looking at statistical trends and patterns that suggest what a stock will do in the future.

key metrics: moving averages, trendlines, and momentum indicators.

THE RISE OF REDDIT



THE RISE OF REDDIT

THANKSGIVING MAKE SURE YOUR FRIENDS DO TO. THE WORST FEELING IN NOT GETTING AN INVITE TO THANKSGIVING OR CHRISTMAS." u/IDELETEDMYACCOUNT#4 "BEN SHAPIRO'S SISTER HAS ABSOLUTE HAMMERS AND I WOULD MURDER MY FAMILY TO HAVE HER SIT ON MY FACE" u/DRSELUSS#1#2# "WHY WOULD I SELL AT A 50% LOSS WHEN I CAN SELL FOR A 95% LOSS LATER" u/DIAMOND-NUTZ "TIRED OF SEEING THAT FTX GUY'S FAT FACE EVERYWHERE." u/SAMOFNY "BEARS GETTING THEIR ASSHOLES STUFFED LIKE A TURKEY ON THANKSGIVING" u/ROOTNP#C "MY PUTS... LOOK WHAT THEY DID TO MY PUTS" u/COMFORTABLEOWL#0 "I'VE LOST \$10,000 IN THE PAST TWO DAYS" u/NOTPAPAELON

wallstreetbets [Join](#)

r/wallstreetbets

Posts Predictions **LIVE** Wiki YouTube Discord Twitter Twitch Thread Filters (NEW) ▾

Hot New Top ...

PINNED BY MODERATORS

40

Posted by u/OPINION_IS_UNPOPULAR AutoModerator's Father 4 hours ago

Daily Discussion Thread for November 23, 2022 [Daily Discussion](#)

2.8k Comments Share Save ...

About Community

Like 4chan found a Bloomberg Terminal

Created Jan 31, 2012

13.1m • 40.4k #25

Degenerates Buying FDs Ranked by Size



Dinglejingle88 🔥 +3 · 4 min. ago

I love getting bent over a barrel on a bid price 😞



Reply

Share

Report

Save

Follow

Comments Per Day

14175

Rank

8



LowerDenominator 🔥 +2 · 4 min. ago

and i sold nvda at 147. fuk me



Reply

Share

Report

Save

Follow



fscumeau 💎 +4 · 5 min. ago



TSLA LMAO



Reply

Share

Report

Save

Follow



ComfortableOwl0 🔥 +2 · 5 min. ago

Eats sidewalk bread

I fuckin give up bro



Reply

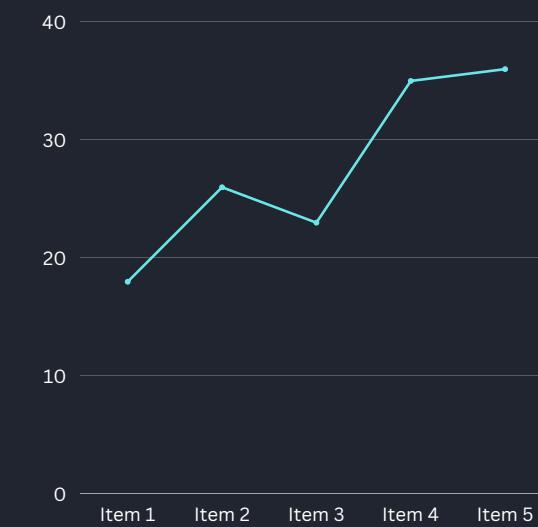
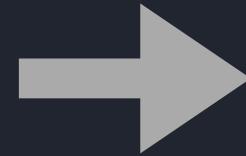
Share

Report

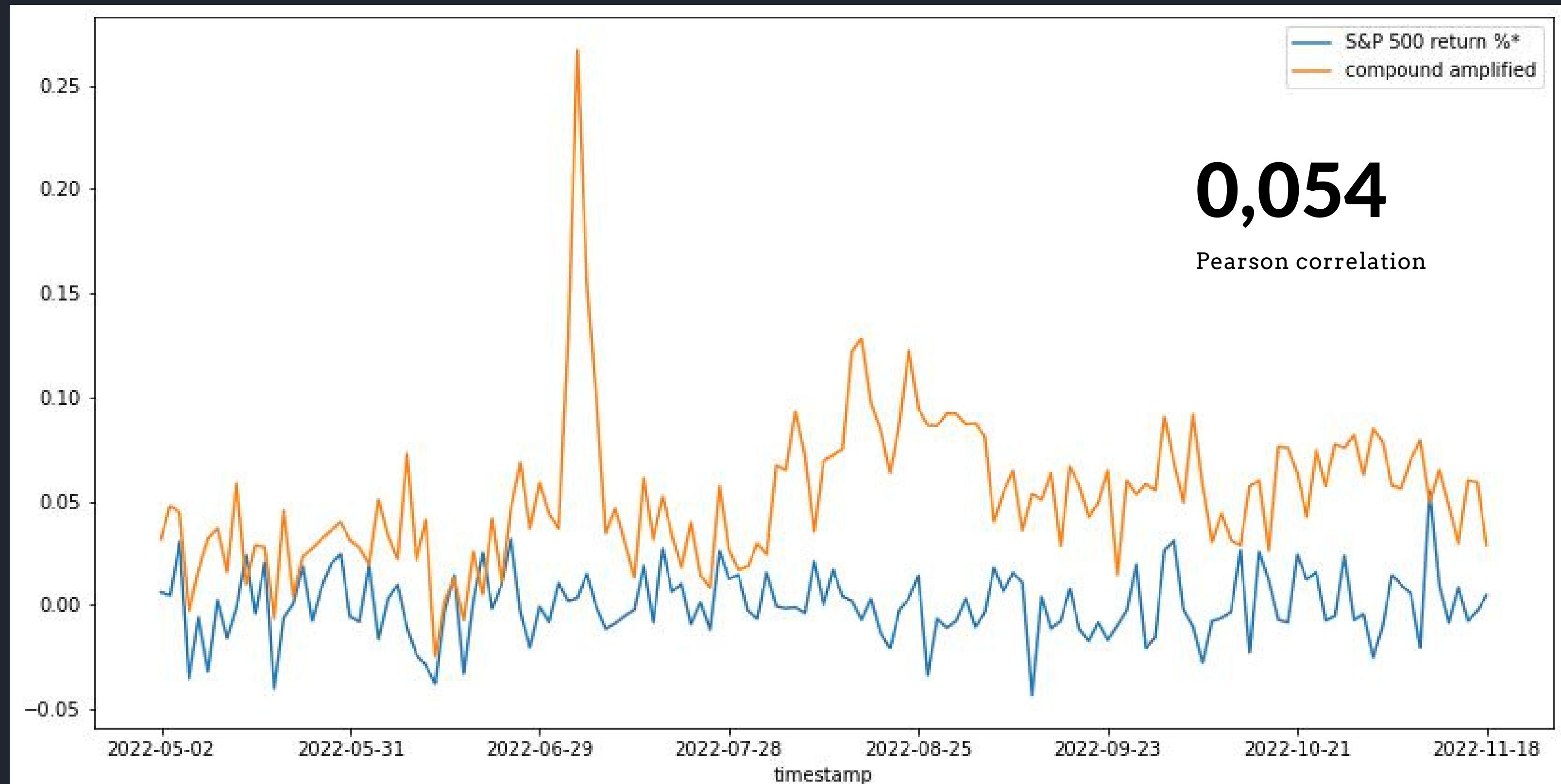
Save

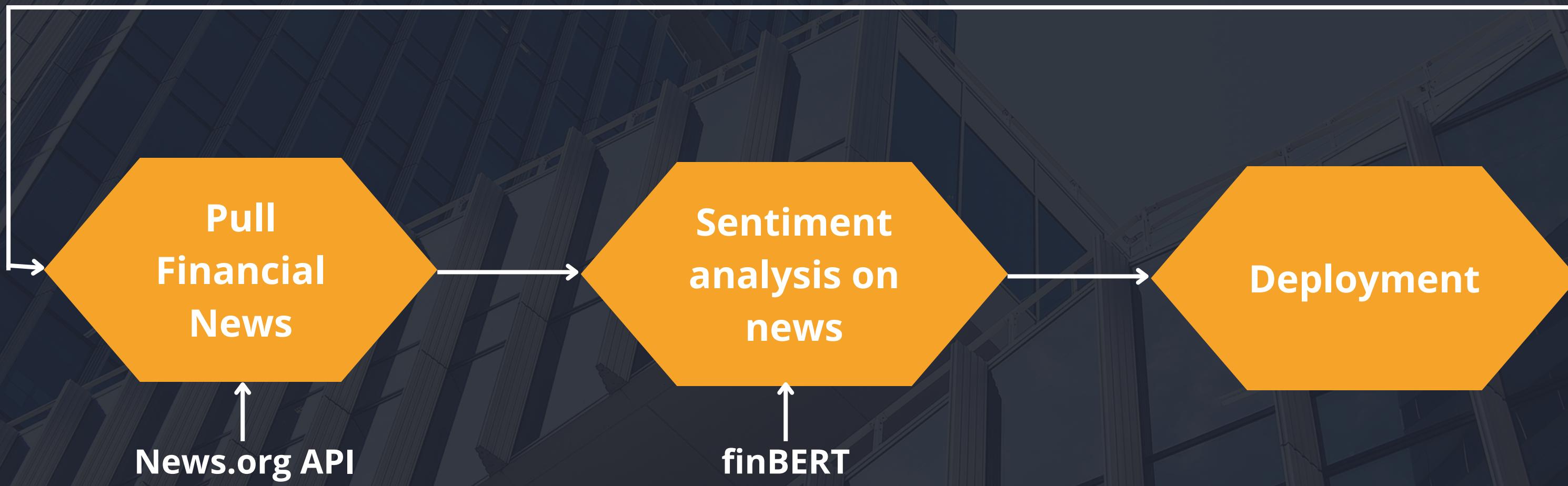
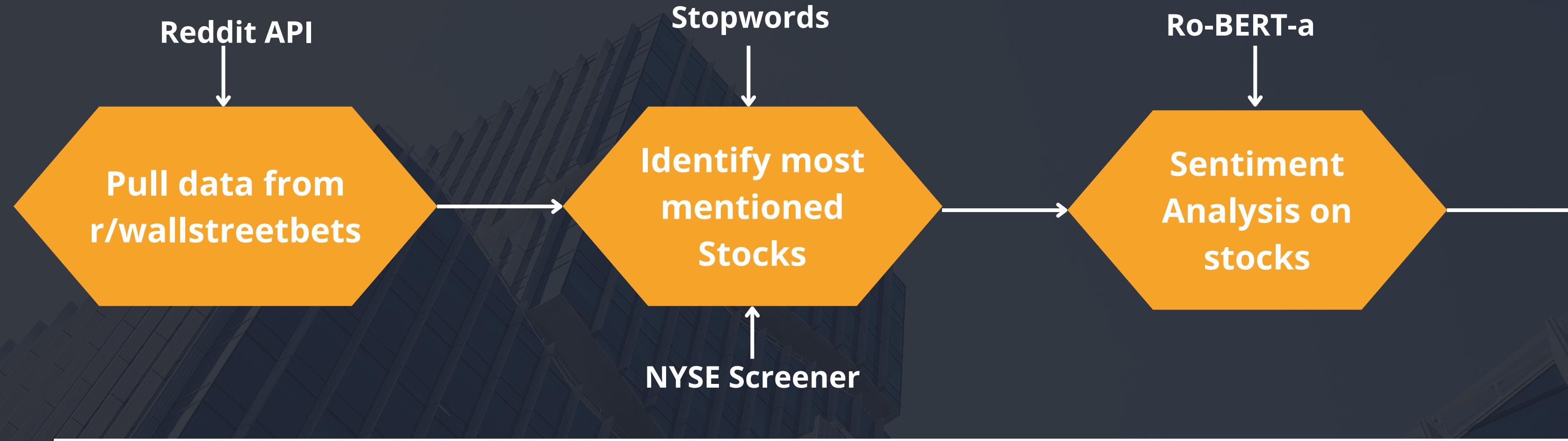
Follow

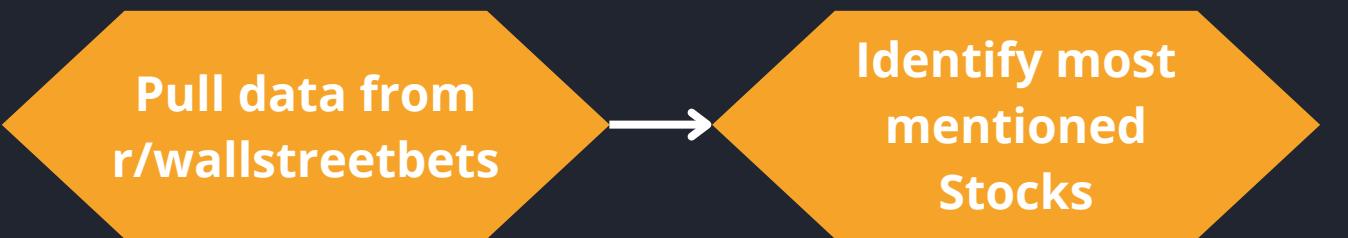
CAN WSB SENTIMENT PREDICT THE MARKET?



WSB SENTIMENT VS. S&P 500

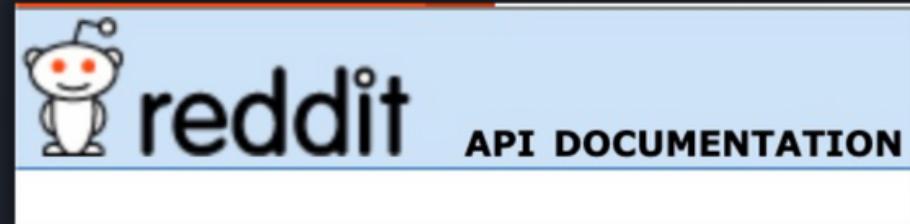




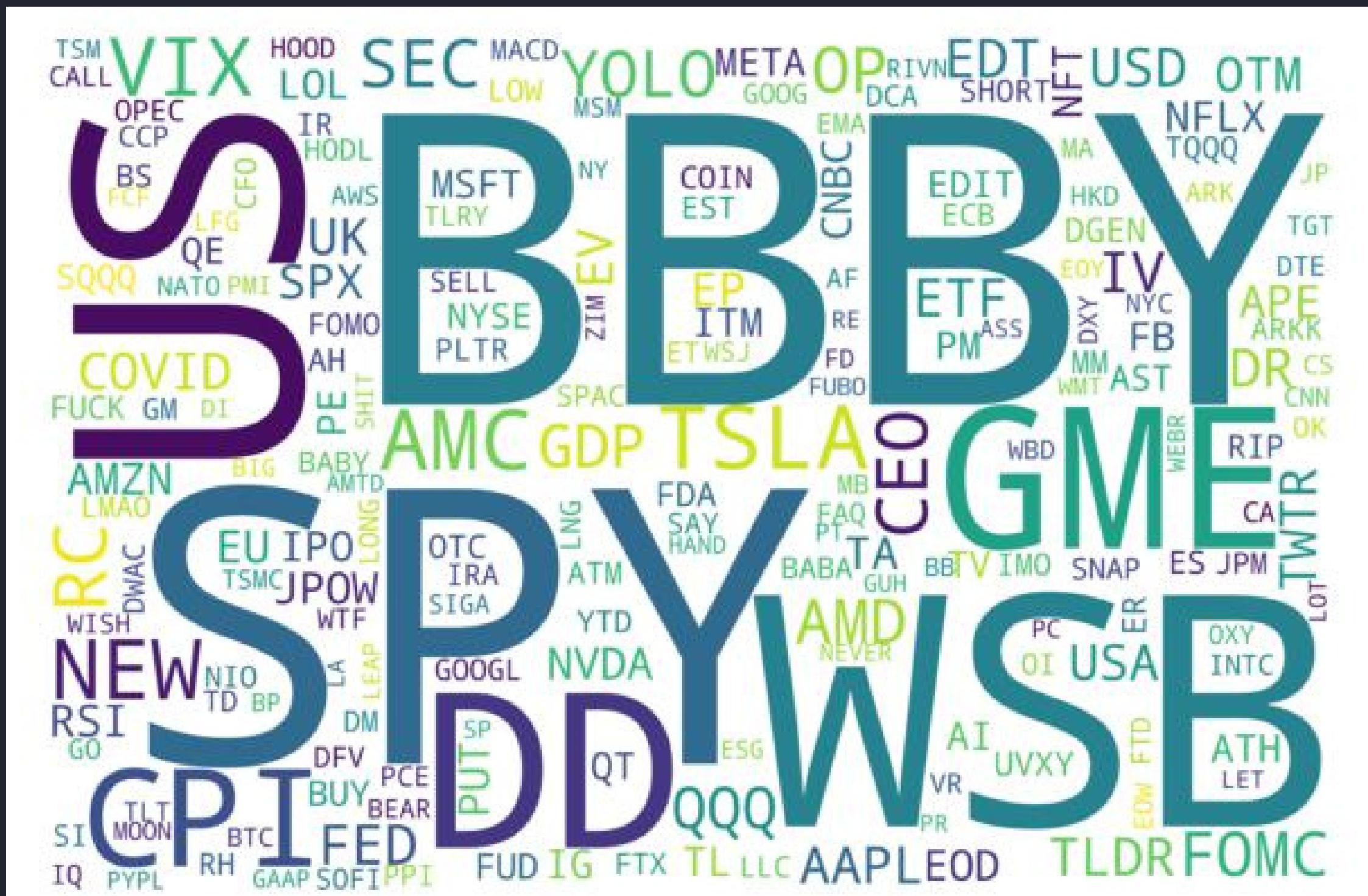


```
sub='wallstreetbets'

def retrieve_reddit_data():
    reddit = cs.auth()
    postList = cs.getThread(reddit,sub)
    commentsListandTime = cs.getComments(postList)
    tickerList = cs.getTicker(postList)
    validList = cs.validateTicker(tickerList)
    df3 = cs.filtering_relevant_stocks(validList)
    df3 = cs.formatting2(df3)
    stocks = list(set(df3['Ticker']))
    comments = list(df3['comment'])
    return df3
df3 = retrieve_reddit_data()
```









BERT

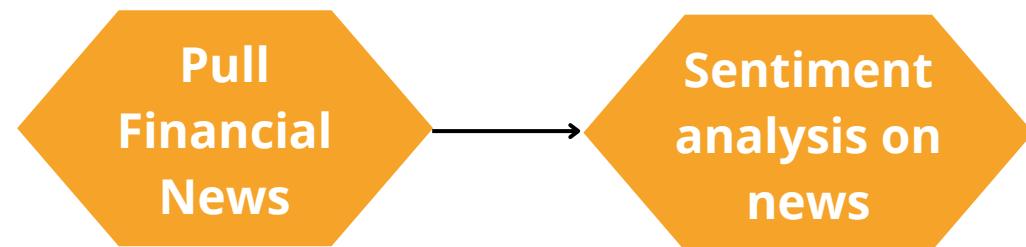
Bidirectional Encoder Representation from Transformers

- **Context into account:** opposed to unidirectional models that evaluates the text from left to right. Bert evaluates all neighbours of each word together
- **Vast knowlegde of words and sentences:** opposed to lexicon/rule models that attributes sentiment based on lexicon of 8k tokens. BERT was trained on millions of different texts through unservised learning
- **Fine-tunning:** Bert can be fine-tunned on smaller datasets for specific tasks --> **roBERTa** (Variation of BERT fine-tunned for twitter data)

CONTEXT IMPORTANCE

GUH

“GUH” is an internet term used to mean someone is annoyed or upset. In the context of WallStreetBets, GUH is likely to be used when someone has lost a lot of money on an investment. For example, someone might use GUH in the title of a Reddit post where they show a stock chart that’s sloping severely downward.



NEWS.API

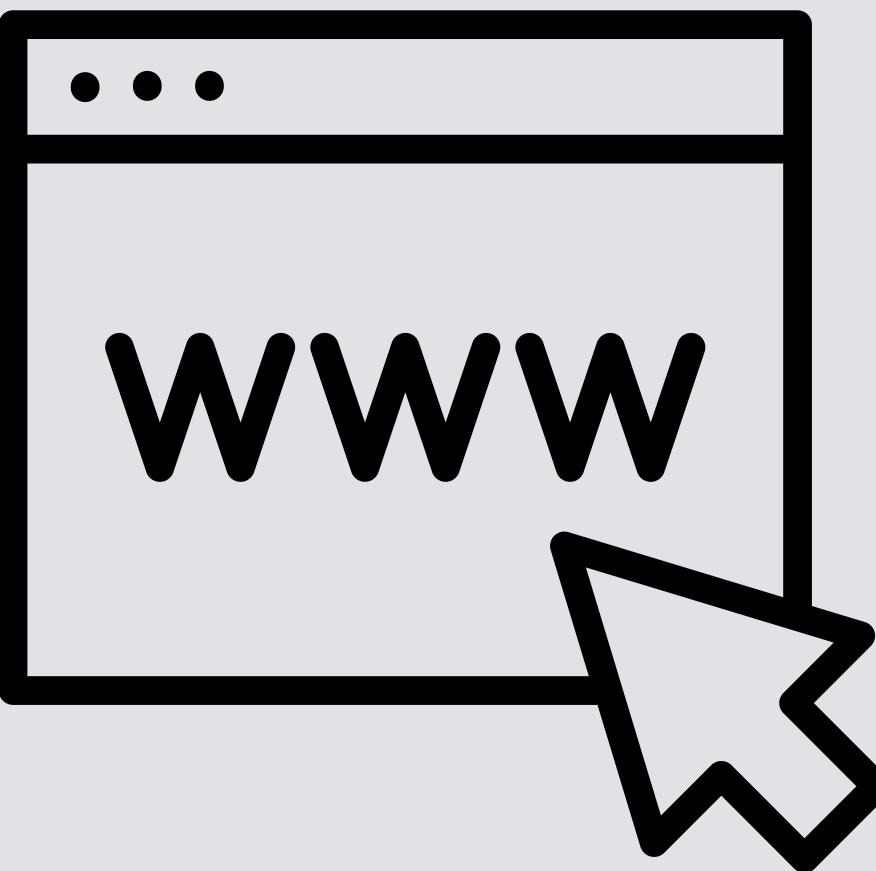
30.000 news sources

including relevant ones for stock market:
Bloomberg, Business Insider, google news,
CNN, Wall Street journal, Washignton Post

```
k = list(zip(stocks, stock_name))
d={}
articles=[]

for (x, y) in k:
    news = newsapi.get_everything(
        q=f'{x} AND {y}',
        language='en',
        from_param=str(yesterday),
        sort_by='relevancy')
    articles = news['articles']
    articles = pd.DataFrame(articles)
    articles = articles['content']
    d[x]=articles
df = pd.DataFrame(d)
```

Deployment



SHORTCOMINGS



01

Reddit's relevance

1 year after WSB Short Squeeze, Reddit's influence is still to be recreated

02

Spotting Momentum

Sentiment on popular stocks is part of the story. Distinguishing them from real trading opportunity would add more value

03

Scalability constraints

APIs' requests limitation
Computational power

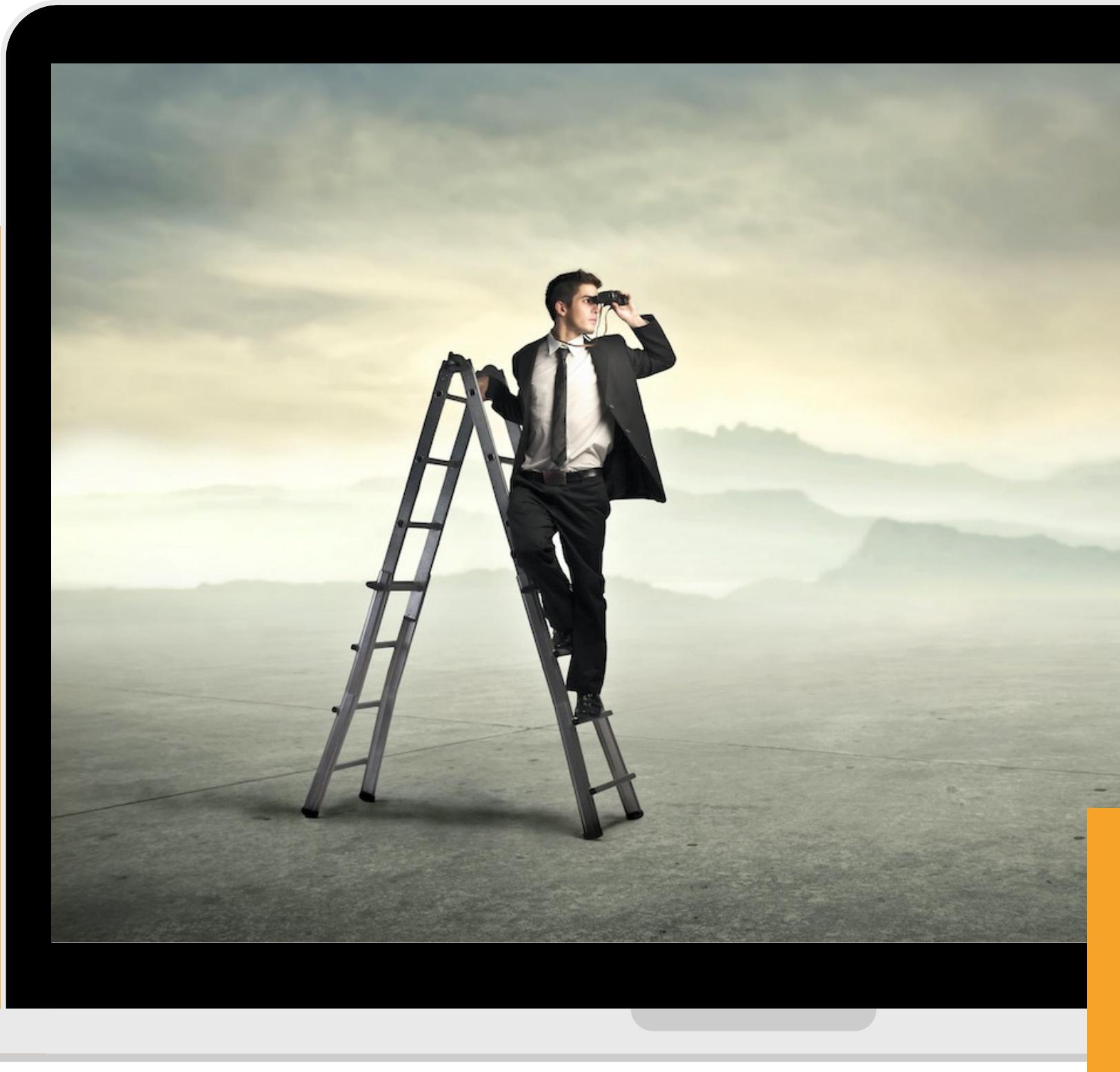
STAY TUNED!

COMING UP NEXT...

Interactivity: giving the user the possibility to look for stock on other subreddits and setting email alerts

Performance enhancements: test a lighter NLP model

New functionalities Additional technical analysis solutions: time series forecast of most mentioned stocks



A large, modern building with a glass and steel facade, viewed from a low angle looking up. The building has a curved, angular design with many windows. The sky is overcast.

THANK YOU