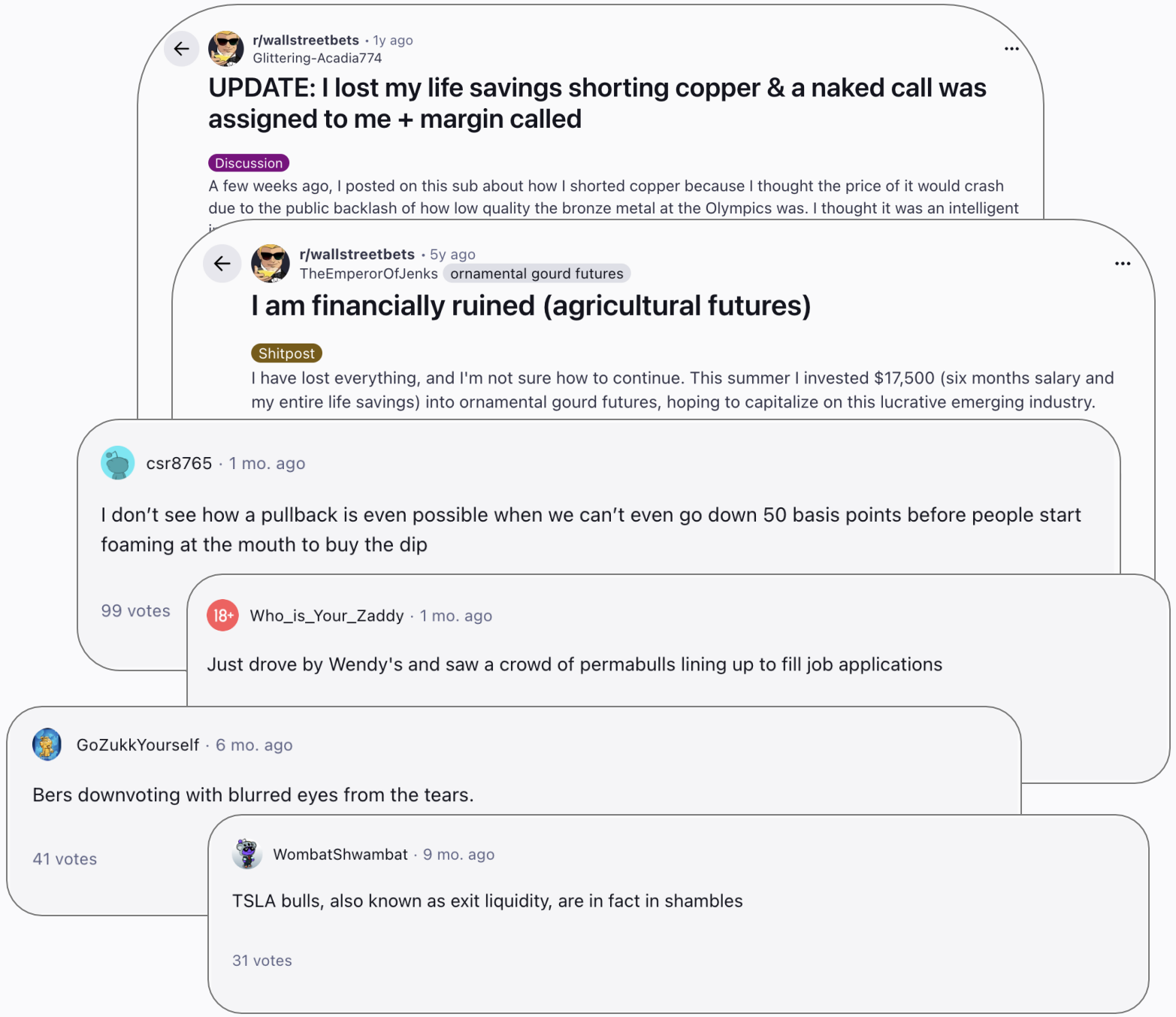
# Can social media be trusted for stock price prediction?

An intraday $SPY trading strategy based on r/wallstreetbets sentiment

*Presented by:*
**Benjamin Luo**



## Section 1 | **Project Overview**

- Motivation
- Project Objectives
- High-Level Architecture

Section 2 | Data Engineering and Scraping

Section 3 | Sentiment Analysis

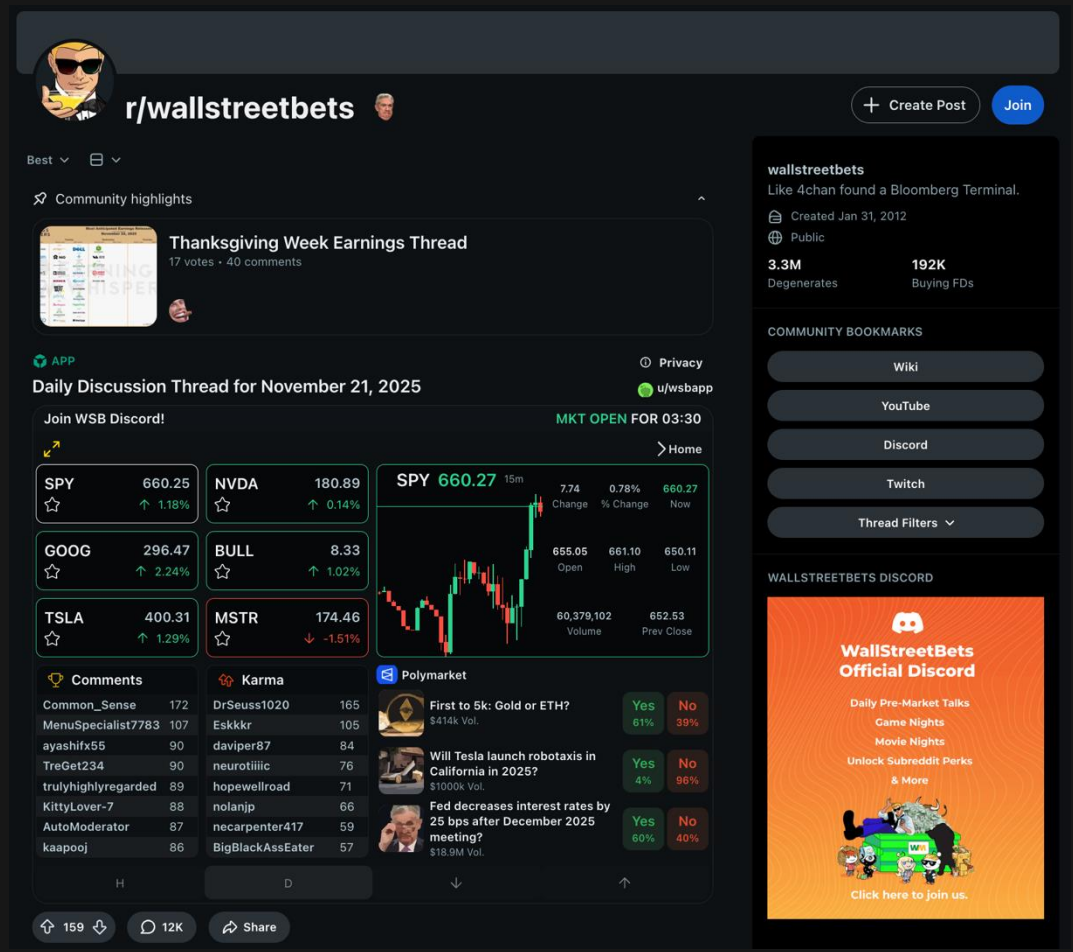
Section 4 | Trading Strategies and Backtesting

Section 5 | Discussion

What is r/wallstreetbets?

# r/wallstreetbets is a subreddit where retail investors discuss high-risk, highly speculative stock and option trading

The r/wallstreetbets **daily discussion thread** can provide **real-time** market news and **sentiment signals**




## Limitations

- **Bots and trolls** create noise and false signals
- **Linguistic complexity** (typos, slang, sarcasm, ambiguity)
- **Low predictive power** (**reactive** sentiment, short-term focus)

reddit.com/r/wallstreetbets/

# The terminology deviates from standard financial language, requiring models to be carefully trained to correctly interpret the true intent

| Phrase | Meaning |
|---|---|
| Ber | Bear; someone who wants the price to go down |
| Bol | Bull; someone who wants the price to go up |
| Diamond hands | Holding onto a position for dear life |
| Paper hands | Closing a position early on |
| To the moon 🚀 | Extreme optimism |
| Hedgies | Hedge funds |
| Bagholder | An investor left holding the “bag” after a price crash |
| Wendy’s | The traditional, unofficial employment location for WSB traders who have lost all their money |
| Regard | A euphemism for “retard”; a bad trader |
| Exit liquidity | Bagholders who are used for liquidity by large investors |
| 0dte | 0 days-to-expiry options |


**BearyChristmas223** • 1mo ago

Things visible from space:

- Great Wall of China
- Giza pyramids
- The bags of BYND shareholders

**Main-Economist67** • 19d ago

I refuse to take any accountability for my trades. All of my losses are due to bad luck, all of my gains are due to skill.

**bullrfuk** • 19d ago

WSB whenever there is the slightest disturbance: "WW3 is coming"

WSB when the disturbance ends: "it was so obvious"



The BEAR after earnings!

**GemmyBoy999** • 28d ago

I'm somewhat of a wall street philanthropist myself

**Visual_Enthusiasm_73** • 17d ago

Quick! Everybody take out loans and lines of credits and let's pump this back up!

**jsie-iaiqhsi816278** • 3mo ago

☝it will be

👈frankly 👉

the greatest👌

🖐economic crash🖐

👌in the

☝history of

👈 amaerica👉

They will say. Mr 🍊 no one can crash it like you 👈👉

**hv876** • 7d ago

Top 1% Commenter

They called you an investor

**---Right--Tackle---** • 18d ago

Top 1% Commenter

At this rate I'm gonna have to return my groceries

**Sweg_OG** • 18d ago

too scared to even open the app anymore

**KittyLover-7** • 6mo ago

Top 1% Commenter

My greed is the greed they warned about in the bible

Disclaimer

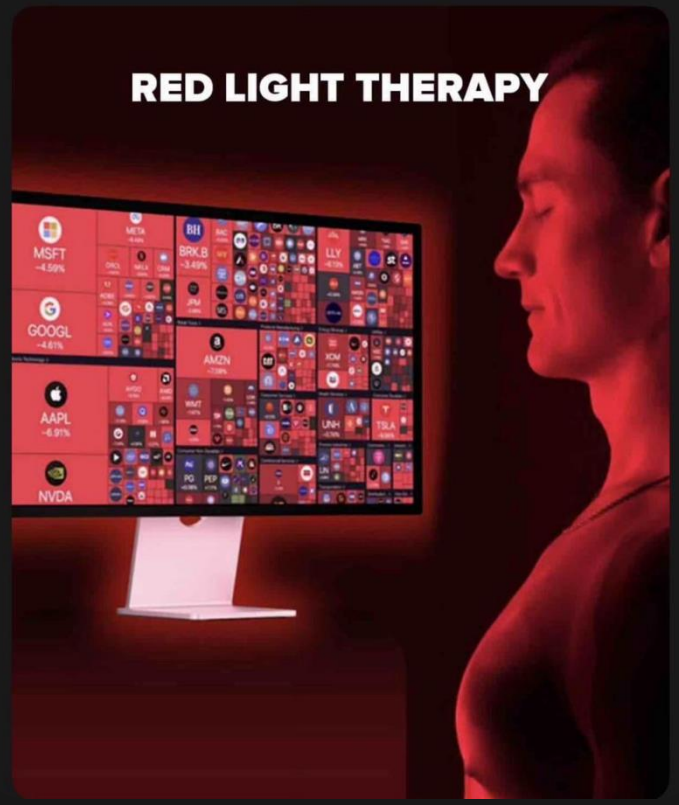
This could be you.

Trade with **caution**!!

Skincare first ladies & gentlemen 💪

Laughing through the pain 🩸🩸🤦🏻‍♀️🤦🏻‍♀️

$VFV $TSLA $MSTR $ASML $GOOGL $NVDA $PLTR

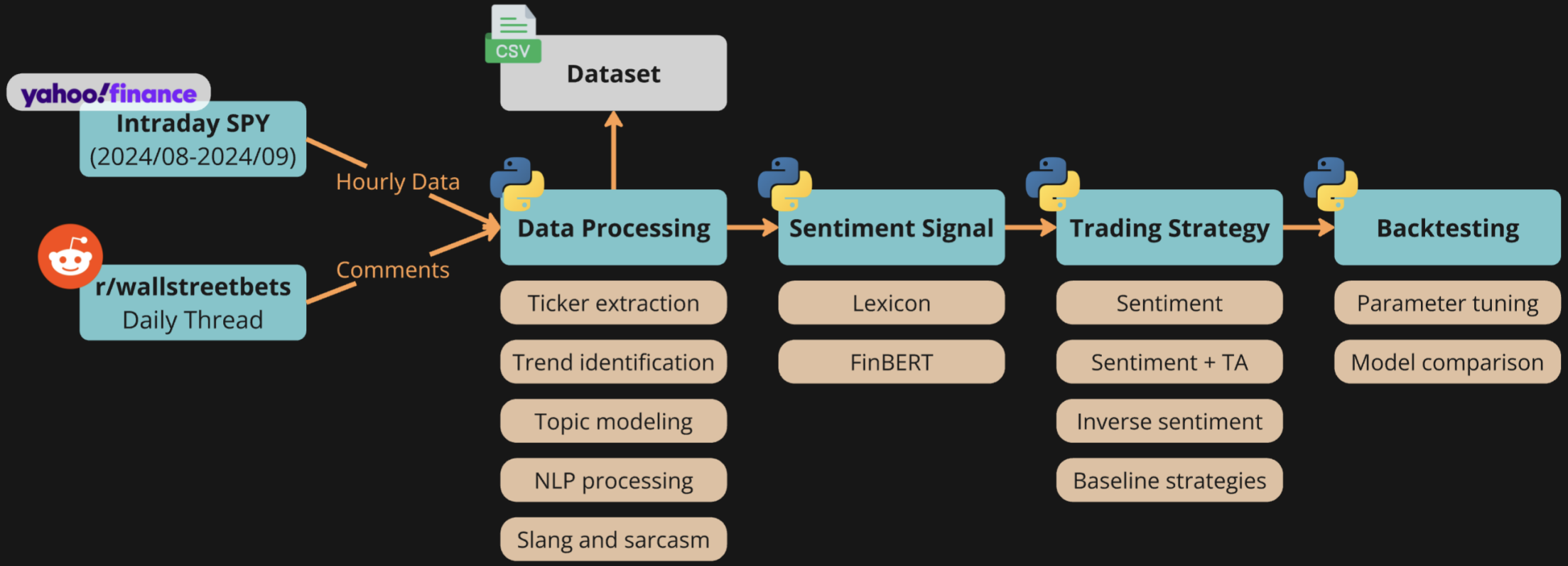





Project Overview

# Develop a trading strategy for intraday $SPY using r/wsb sentiment

Yahoo Finance and Reddit APIs provide historical and real-time data for algorithmic trading



Section 1 | Project Overview

## Section 2 | **Data Engineering and Scraping**

- Yahoo Finance API for Market Data
- Reddit API (PRAW) for r/wallstreetbets Data
- Ticker Extraction
- Trend Identification
- Topic Modeling
- NLP Processing

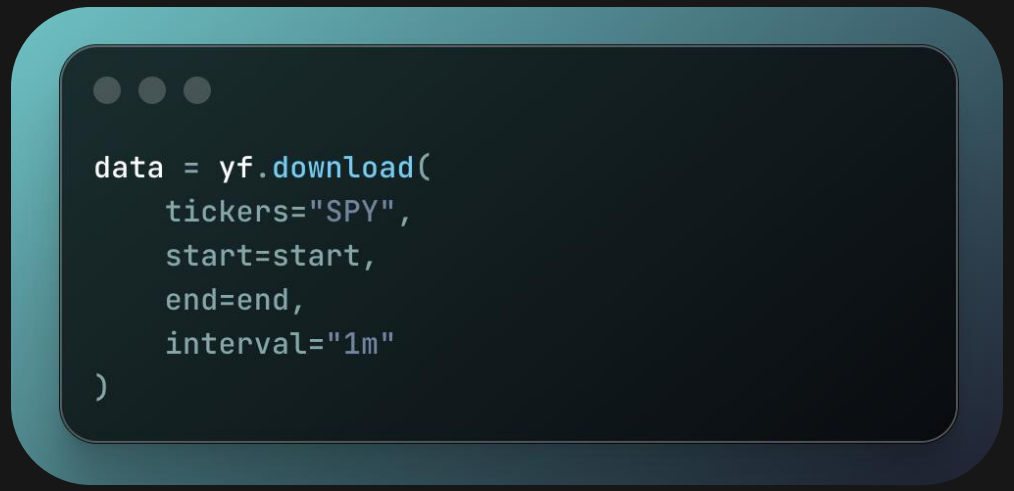
Section 3 | Sentiment Analysis

Section 4 | Trading Strategies and Backtesting

Section 5 | Discussion

# Yahoo Finance provides API access to intraday $SPY OHLC data

I initially used sourced market data from **Kaggle**, but realized **yfinance** is better for **on-demand** and **real-time data**



```
data = yf.download(
    tickers="SPY",
    start=start,
    end=end,
    interval="1m"
)
```

**Yahoo Finance API**

Start and end dates are set based on the comment dataset

| timestamp | open | high | low | close | volume |
|---|---|---|---|---|---|
| 2024-08-19 4:00 | 559.54 | 559.54 | 553.98 | 554.15 | 523 |
| 2024-08-19 4:01 | 554.15 | 554.22 | 554.05 | 554.05 | 982 |
| 2024-08-19 4:02 | 554.07 | 554.09 | 553.95 | 553.95 | 1718 |
| 2024-08-19 4:03 | 553.95 | 553.98 | 553.68 | 553.68 | 1151 |
| 2024-08-19 4:04 | 553.68 | 553.79 | 553.65 | 553.77 | 5232 |
| 2024-08-19 4:05 | 553.7 | 553.74 | 553.3 | 553.3 | 1374 |
| 2024-08-19 4:06 | 553.29 | 553.36 | 553.19 | 553.31 | 114 |
| 2024-08-19 4:07 | 553.27 | 553.33 | 553.12 | 553.15 | 4268 |
| 2024-08-19 4:08 | 553.17 | 553.36 | 553.16 | 553.32 | 155 |
| 2024-08-19 4:09 | 553.25 | 553.5 | 553.25 | 553.5 | 1321 |
| 2024-08-19 4:10 | 553.5 | 553.61 | 553.49 | 553.58 | 1057 |
| 2024-08-19 4:11 | 553.64 | 553.66 | 553.5 | 553.61 | 1529 |
| 2024-08-19 4:12 | 553.61 | 553.66 | 553.56 | 553.66 | 213 |
| 2024-08-19 4:13 | 553.64 | 553.67 | 553.57 | 553.67 | 594 |
| 2024-08-19 4:14 | 553.63 | 553.63 | 553.5 | 553.5 | 29 |
| 2024-08-19 4:15 | 553.59 | 553.76 | 553.55 | 553.76 | 579 |
| 2024-08-19 4:16 | 553.71 | 553.92 | 553.71 | 553.82 | 133 |
| 2024-08-19 4:17 | 553.84 | 553.84 | 553.84 | 553.84 | 55 |
| 2024-08-19 4:18 | 553.79 | 553.88 | 553.71 | 553.71 | 1050 |

**Intraday $SPY ETF**
(1 min)

# PRAW (Python Reddit API Wrapper) provides r/wallstreetbets comments data

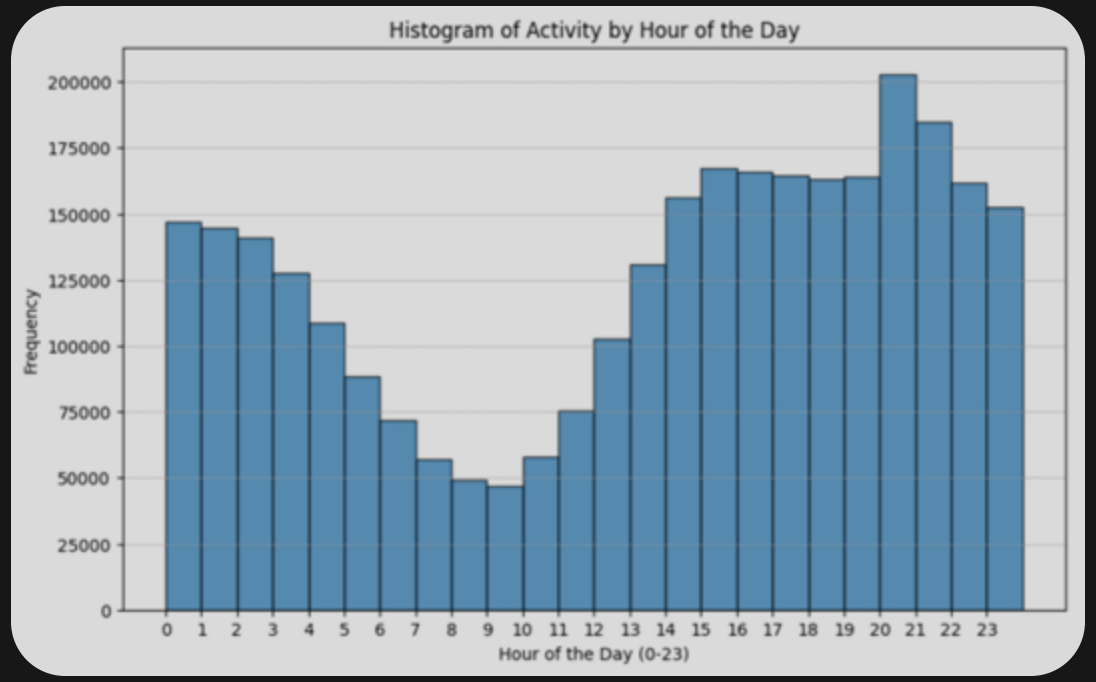
It takes ~20-30 minutes to retrieve comments from a daily thread (~11k comments) due to rate limits, so the time horizon is limited

| id | author | score | created_utc | created_datetime | body |
|---|---|---|---|---|---|
| nr7ede4 | FilipMKD | 73 | 1764331450 | 2025-11-28 7:04 | The  market can't be red on black Friday. Can't be scared when buying that air fryer |
| nr7ebhq | Jesus_Right_Nut | 64 | 1764331423 | 2025-11-28 7:03 | Due to my portfolio's health, I too have decided to pardon a turkey this year |
| nr7hhad | FabulousLanguage6216 | 59 | 1764332998 | 2025-11-28 7:29 | You watch The big short I watch The wolf of wall street.  we are not the same |
| nr7emf5 | DwigtSchrute1 | 56 | 1764331579 | 2025-11-28 7:06 | Put the servers outside, it,Äôs cold as shit in Chicago |
| nr7f2l2 | yaboisthename | 44 | 1764331809 | 2025-11-28 7:10 | I'm so glad I got to experience black friday back when it was metal people getting trampled n shit |
| nr7q318 | Internal_Field5970 | 40 | 1764336715 | 2025-11-28 8:31 | There,Äôs a liquidation event happening in my bathroom¬† |
| nr7sco7 | aftherith | 39 | 1764337606 | 2025-11-28 8:46 | One month left to REALLY embarrass yourself in front of your accountant |
| nr7evv0 | DahyunDabFan | 42 | 1764331714 | 2025-11-28 7:08 | CME shut down cuz my calls were getting too hot |
| nr7fw5h | Outof_ITM | 32 | 1764332220 | 2025-11-28 7:17 | Selling dick.  $40/inch or get all three for $100 |
| nr7ixi9 | callsonreddit | 33 | 1764333689 | 2025-11-28 7:41 | I need to make 5k today so I can buy something on black friday for my cat |
| nr7yadu | Chicagosjuice | 30 | 1764339862 | 2025-11-28 9:24 | Unacceptable. How can I sue the stock market? |
| nr7ejnb | shadowban1244 | 25 | 1764331540 | 2025-11-28 7:05 | As much as I wanna play today, I think imma take my regard cap off and wait for Monday |
| nr7fnc9 | Me55y | 28 | 1764332098 | 2025-11-28 7:14 | How long does it take to restart some fuckin computers |
| nr7rlv8 | aftherith | 23 | 1764337316 | 2025-11-28 8:41 | Some of y'all just need 200-300% per day to break even in the year. Keep going □üí™ |
| nr7i6qz | grimandnordic1 | 20 | 1764333341 | 2025-11-28 7:35 | That massive shit you take the morning after Thanksgiving □üò´□üçó |
| nr7haie | L2F_mens_glutes | 18 | 1764332907 | 2025-11-28 7:28 | #real men marry single moms with 4 kids from different men  Real men of genius |
| nr8hs0g | Yousoldmetohigh | 20 | 1764346517 | 2025-11-28 11:15 | I regret these weekly nvidia calls. |
| nr7ipni | Maximumm_Drawdowns | 18 | 1764333588 | 2025-11-28 7:39 | Remember friends - the casino will be closing early today (1pm est) |

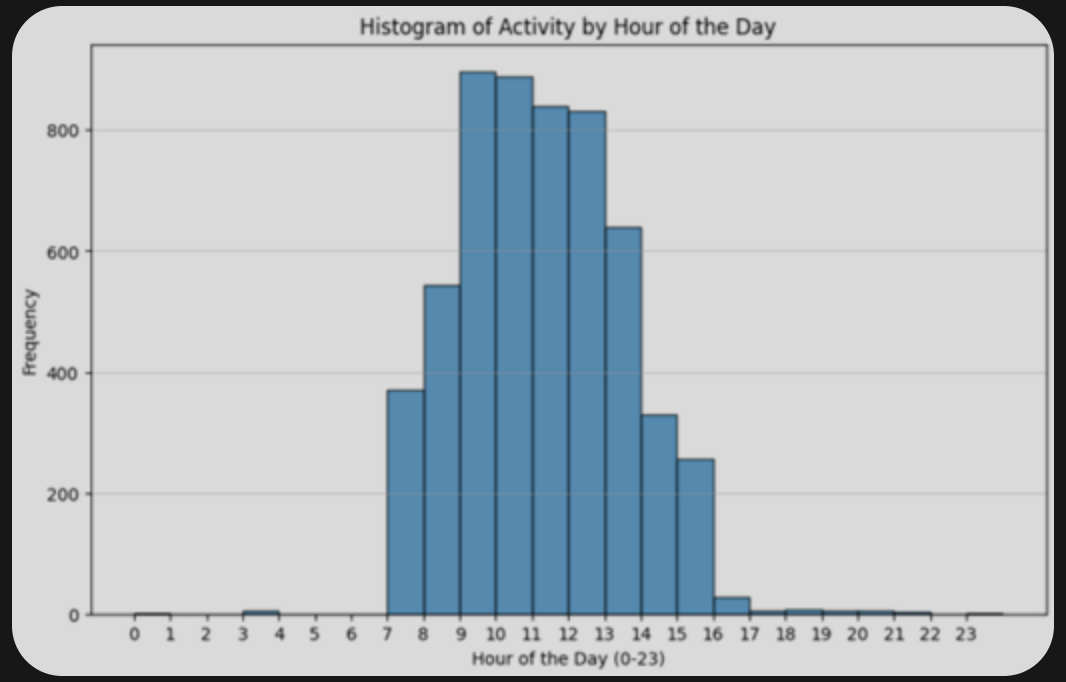
**r/wallstreetbets Data**
(Daily Discussion Thread)
n ≈ 148k (19 trading days)

# Kaggle also provides r/wallstreetbets comment data but contains lots of missing data

Kaggle was initially used but it produced poor results for intraday trading due to data sparsity, albeit covering a greater time horizon




**Kaggle dataset**
(2023-2025)




**Manually extracted dataset**
(25/11/28 daily discussion)

r/WallStreetBets Posts & Comments. https://www.kaggle.com/datasets/curiel/rwallstreetbets-posts-and-comments

# Extracting tickers helps reduce noise and enables the development of stock-specific strategies

**Spacy's** pretrained en_core_web_lg model was used to extract tickers

Untitled-1

```python
def ticker_extraction(text):
    """Extract the ticker symbol(s)"""
    doc = nlp(str(text))
    tickers = [ent.text for ent in doc.ents if ent.label_ == "ORG"]
    unique_tickers = list(set(tickers))

    return ", ".join(unique_tickers) if unique_tickers else None
```

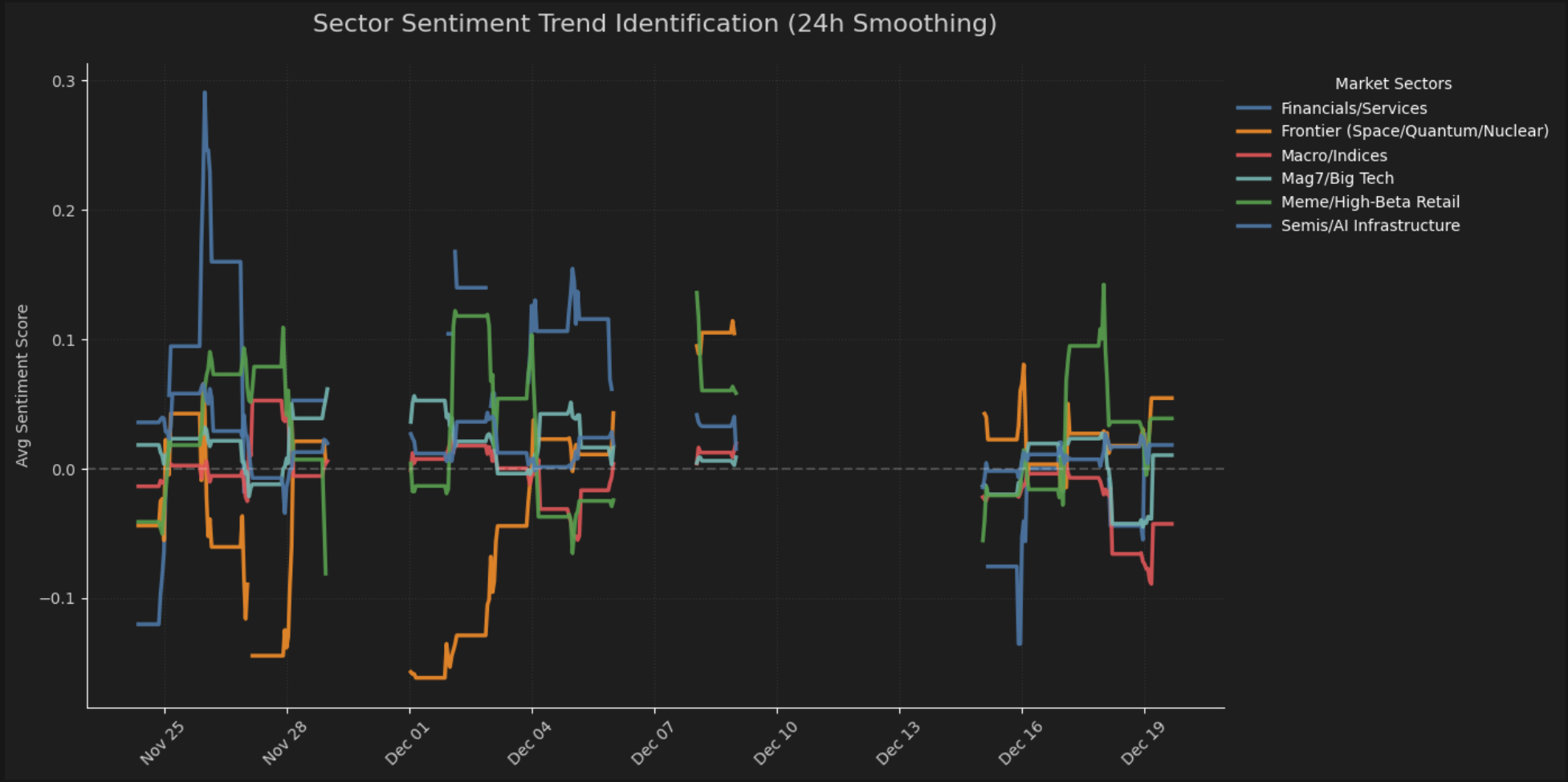
**Code for extracting tickers**

| text | ticker |
|---|---|
| Who the fuck is dumping GOOGL | GOOGL |
| Can,Äôt believe I,Äôm wasting my twenties l | |
| We haven't had a Santa rally in the last 5 fuck | |
| options i was looking at but didn't enter: +130 | |
| $44,000 gain in 30 seconds scalping $META | META |
| META's AI problems erased by cutting 30% o | Metaverse, META |
| For the love of God stop looking at the mark | |
| Nvidia bout to take the next gap up and take | Nvidia |
| I really wish there was any other way to make | |
| wtf bro fuck this I,Äôm gonna go jack off I gu | |
| Hi, I,Äôm Sam Altman, I need 7 trillion dollars | |
| https://preview.redd.it/7umixk2rd85g1.png?\ | SPX |
| Yall be tax loss harvesting while I,Äôm tax los | |
| So who shorted the bottom of that dip lmao | |
| Just buy the DIP, bro you should have made a | |
| every week I halve my port. nice. | |
| NVDA PAMP IT | NVDA |

**Sample ticker extraction**

# Tickers can be categorized into sectors and coupled with sentiments to track trends

The graph tracks the average sentiment scores across various market sectors, highlighting significant volatility in the Financials and Frontier sectors compared to the more stable Big Tech and Semis categories



# Word clouds visually represent frequent words in comments with positive or negative sentiment

Comments were disaggregated on sentiment (+1, 0, or -1), key words were identified, then word clouds were created based on word frequency


**Prevalent words with positive sentiment**


**Prevalent words with negative sentiment**

**Methodology from**: WallStreetBets Reddit Posts Analysis. https://www.kaggle.com/code/gpreda/wallstreetbets-reddit-posts-analysis#Data-visualization

# Text cleaning was done using regular expressions, stop words, and tokenization

Stemming/lemmatization were avoided for VADER as its lexicon map contains full-words, and cleaning was unnecessary for transformer-based models (ex. FinBERT) as they are trained on full sentences

```python
def clean_text(text):
    """Data cleaning"""
    text = text.lower()

    # Patterns
    text = re.sub(r'http\S+|www\S+|https\S+', '', text, flags=re.MULTILINE)
    text = re.sub(r'img\s*emote\s*t5\s*2th52\s*\d+', '', text)
    text = re.sub(r'img\s*emote\s*t5\s*\d+', '', text)
    text = re.sub(r'img\s*emote', '', text)
    exclude = set(string.punctuation.replace('!', '').replace('?', ''))
    text = ''.join(ch for ch in text if ch not in exclude)
    text = re.sub(r'\d+', '', text)

    # Stop words
    stop_words = set(stopwords.words('english'))
    wsb_noise = {'im', 'isnt', 'its', 'youre',
                 'theyre', 'dont', 'get', 'like', 'would', 'can'}
    stop_words.update(wsb_noise)
    word_tokens = nltk.word_tokenize(text)
    filtered_words = [w for w in word_tokens if not w in stop_words]

    text = ' '.join(filtered_words)
    text = re.sub(r'\s+', ' ', text).strip()

    return text
```

**Code for cleaning text**

| text | cleaned_text |
|---|---|
| I feel like we are going to find out | feel going find |
| raging bull day in, day out | raging bull day day |
| We break 185 in a few minutes and run | break minutes run |
| When is he talking? | talking ? |
| Special operation comrade | special operation comrade |
| Nobody knows, but my guess is it'll be something minute a | nobody knows guess itll something minute stupid dead kids r |
| probably something on immigration | probably something immigration |
| Gold toilets | gold toilets |
| trust the process | trust process |
| Winners need losers, stop trying to educate the retards wh | winners need losers stop trying educate retards think tpus rep |
| Putin drama | putin drama |
| You don't need Bloomberg. The market is re adjusting bc c | need bloomberg market adjusting bc wanton speculation fun |
| Your mom CUH | mom cuh |
| AVGO Gaped You | avgo gaped |
| Here is your warning to buy calls before it goes up | warning buy calls goes |
| You wish. Retard | wish retard |
| 2 est | est |
| Same □üò≠ | □üò≠ |
| Amazon killed nvda | amazon killed nvda |
| What chip? Like the potato chip? | chip ? potato chip ? |
| Yup. Good news = dump. Bad news = go up | yup good news dump bad news go |
| □üò∂,Äç□üå´ÔÏè | □üò∂,Äç□üå´ÔÏè |
| SPY can't break 684. Pass it on. | spy cant break pass |
| Not much. What's going on with you, sweet cheeks? □üíã | much whats going sweet cheeks ? □üíã |
| With good stop loss set up it worked out ok | good stop loss set worked ok |

**Sample cleaned text**

Section 1 | Project Overview

Section 2 | Data Engineering and Scraping

## Section 3 | **Sentiment Analysis**

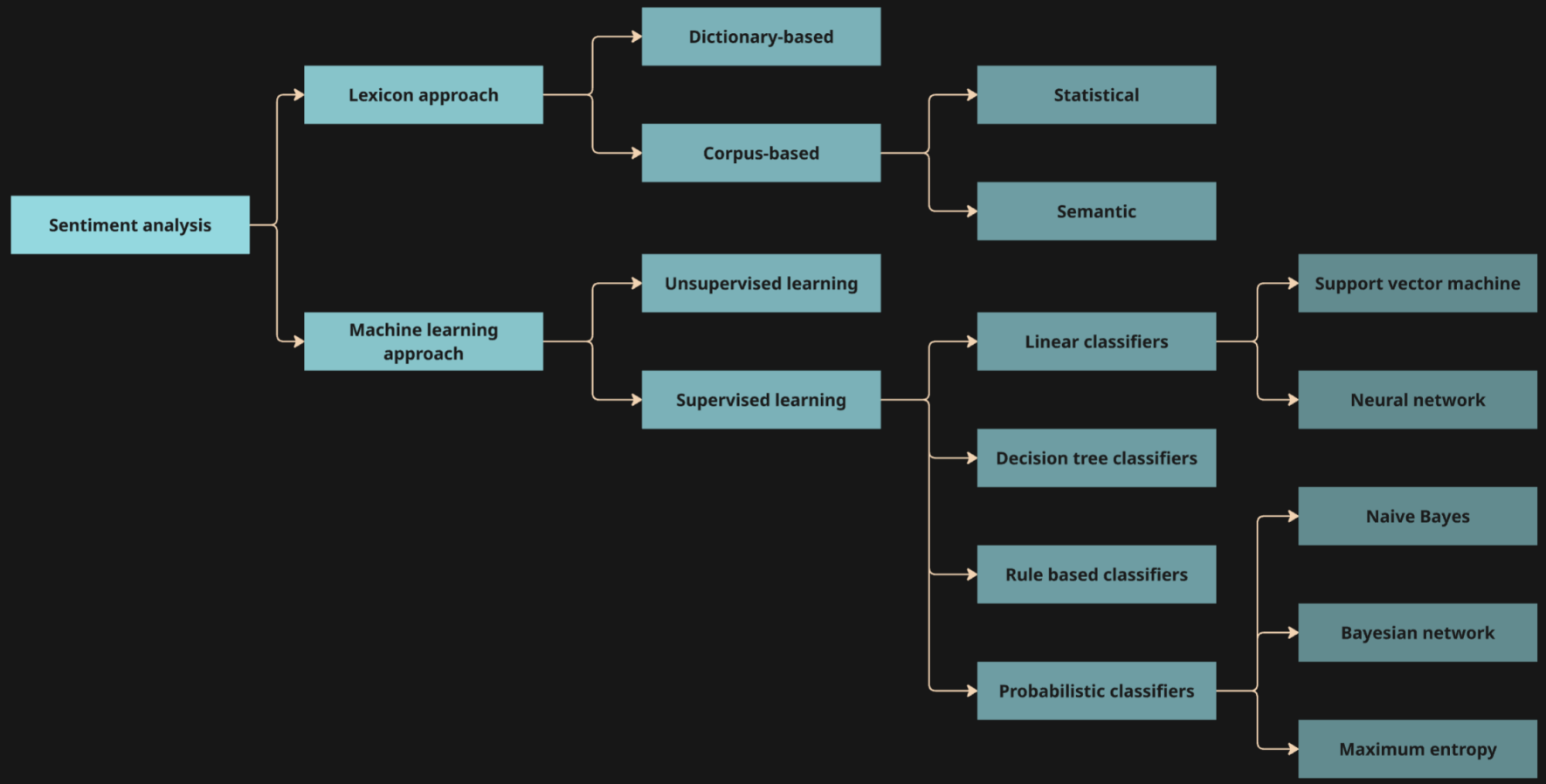
- Approaches to Sentiment Analysis
- Lexicon-based Approaches
- FinBERT
- Data Alignment
- Visualizing Sentiment

Section 4 | Trading Strategies and Backtesting

Section 5 | Discussion

# Sentiment analysis methods are categorized into lexicon and machine learning approaches

Current state of the field research focuses on new context-aware **transformer models**




Sharma, D., Sabharwal, M., Goyal, V., & Vij, M. (2020). Sentiment analysis techniques for social media data: A review. In A. K. Singh, M. K. Sharma, M. Sabharwal, & M. S. W. P. L. Singh (Eds.), Machine learning and cognitive computing for mobile big data (pp. 75–90). Springer. https://doi.org/10.1007/978-981-15-0029-9_7

# Sentiment was estimated by various models, including lexicon-based and ML-based

| cleaned_text | score_vader | score_lexicon | score_financial_lex | score_finbert | ensemble_sentiment | nltk_sentiment |
|---|---|---|---|---|---|---|
| put holder save money enjoy evening | 0.7506 | 0 | -0.05 | 0.052458115 | 0.188264529 | 1 |
| yes maybe also columbia mexico couple countries | 0.4019 | 0 | 0 | 0.039278762 | 0.110294691 | 1 |
| splatter fomo | 0 | 0 | 0 | -0.052418783 | -0.013104696 | 0 |
| announcement still coming ,Äô happened yet | 0 | 0 | 0 | -0.086799111 | -0.021699778 | 0 |
| nah announcement happening minutes theory separate cabi | -0.1027 | 0 | 0 | -0.051634725 | -0.038583681 | -1 |
| karma ponzi demogorgons | 0 | 0 | 0 | -0.128532059 | -0.032133015 | 0 |
| post video little ago | 0 | 0 | 0 | 0.000485189 | 0.000121297 | 0 |
| try hard kind dude type | 0.4588 | 0 | 0 | 0.07595198 | 0.133687995 | 1 |
| volume low much push either direction | -0.2732 | 0 | 0 | 0.015507393 | -0.064423152 | -1 |

**Lexicon-based**

- **VADER**: calculates sentiment based on a list of pre-defined words, then uses rules to handle grammar
- **NLTK**: bins VADER scores into [-1, 0, +1]
- **Simple lexicon**: counts "bullish" versus "bearish" terms, incrementing or decrementing the score by 0.5 for every match
- **Financial lexicon**: uses finance-specific terms, especially from r/wallstreetbets

**ML-based**

- **FinBERT**: pre-trained transformer model on financial news to understand industry-specific context
- **Ensemble**: averages the signal from FinBERT and the lexicon-based approaches

Data Alignment

# The finance and r/wsb datasets were merged on 60-minute time intervals

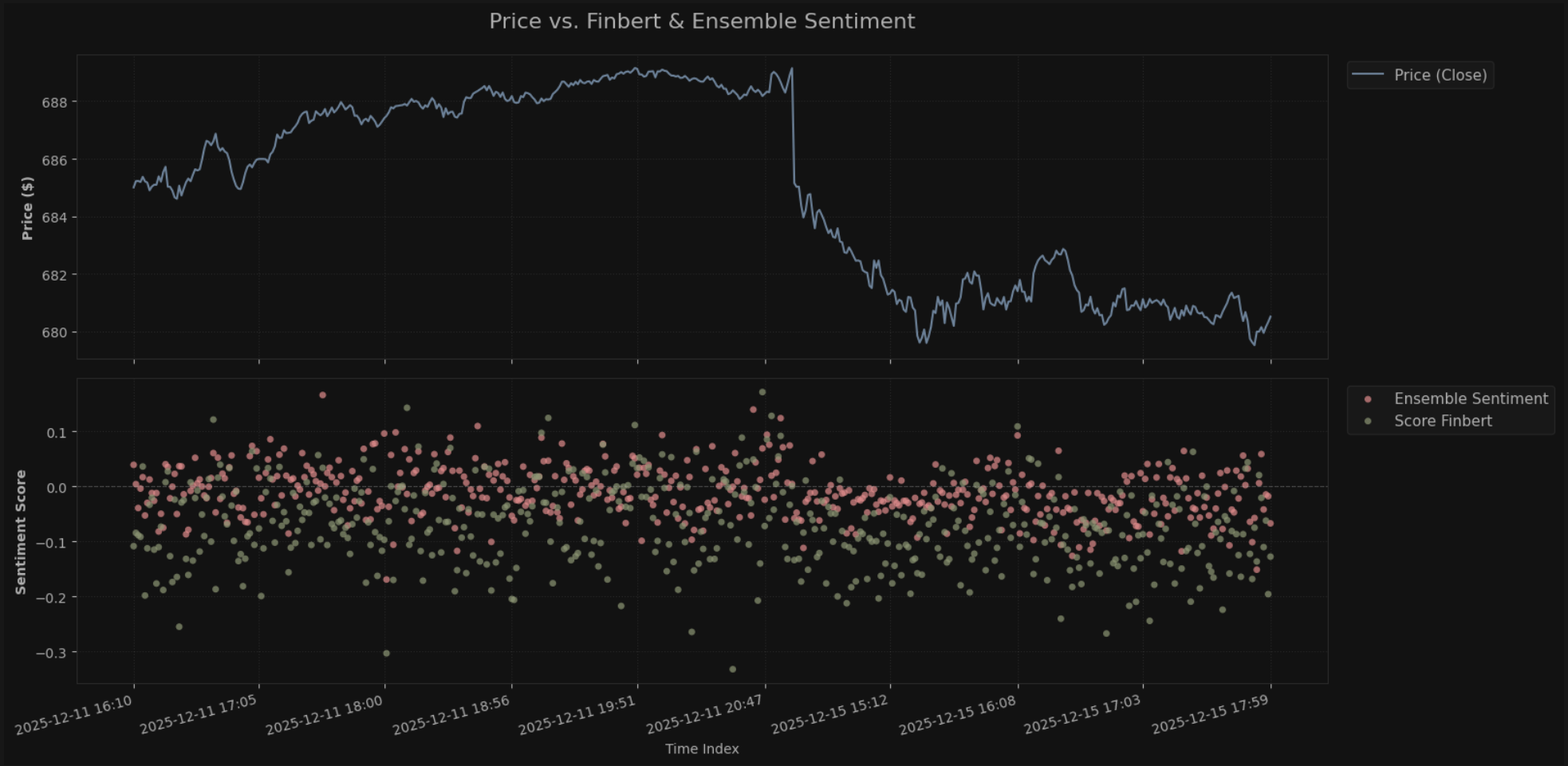
Once grouped by the minute, there was an average of 16 comments per minute, which is fairly sparse

| timestamp | open | high | low | close | volume | upvotes | count | ensemble | finbert |
|---|---|---|---|---|---|---|---|---|---|
| 2025-11-24 14:30:00+00:00 | 662.69 | 663.11 | 662.38 | 663.05 | 2547785 | 29 | 13 | 0.2637649 | -0.016319 |
| 2025-11-24 14:31:00+00:00 | 663.06 | 663.93 | 663.005 | 663.68 | 374231 | 24 | 18 | 0.0556006 | 0.0674185 |
| 2025-11-24 14:32:00+00:00 | 663.69 | 664.55 | 663.63 | 664.53 | 340748 | 34 | 17 | 0.3145684 | 0.2398229 |
| 2025-11-24 14:33:00+00:00 | 664.54 | 664.67 | 664.34 | 664.575 | 382065 | 43 | 19 | 0.534222 | 0.3232029 |
| 2025-11-24 14:34:00+00:00 | 664.59 | 665.07 | 664.38 | 665.0258 | 450566 | 48 | 20 | 0.3612076 | -0.045441 |
| 2025-11-24 14:35:00+00:00 | 665.04 | 665.53 | 665.03 | 665.26 | 497261 | 69 | 33 | 0.5140455 | 0.2844319 |
| 2025-11-24 14:36:00+00:00 | 665.27 | 665.39 | 665.01 | 665.16 | 315633 | 87 | 37 | 0.1659387 | 0.1662664 |
| 2025-11-24 14:37:00+00:00 | 665.17 | 665.31 | 664.39 | 664.56 | 375234 | 76 | 27 | 0.029928 | -0.023995 |
| 2025-11-24 14:38:00+00:00 | 664.56 | 664.67 | 664.18 | 664.23 | 281699 | 32 | 14 | 0.1688871 | 0.3020353 |
| 2025-11-24 14:39:00+00:00 | 664.22 | 664.26 | 663.77 | 663.84 | 277899 | 48 | 26 | -0.269558 | -0.206979 |
| 2025-11-24 14:40:00+00:00 | 663.86 | 664.26 | 663.31 | 663.47 | 352411 | 91 | 32 | 0.1010597 | 0.1757924 |
| 2025-11-24 14:41:00+00:00 | 663.48 | 663.93 | 663.26 | 663.92 | 279330 | 74 | 31 | 0.1788014 | -0.151126 |
| 2025-11-24 14:42:00+00:00 | 663.92 | 664.38 | 663.78 | 664.3 | 259612 | 93 | 29 | -0.055868 | 0.0407053 |

**Merged dataset**

# Sentiment can be visually compared against price movements

High noise is biasing the sentiment signal, but low volume (~16 comments/min) makes filtering risky



Section 1 | Project Overview

Section 2 | Data Engineering and Scraping

Section 3 | Sentiment Analysis

## Section 4 | **Trading Strategies and Backtesting**

- Trading Strategy Parameters
- Baseline Strategies
- Sentiment-based Strategies
- Strategy Comparison

Section 5 | Discussion

# Backtesting was done using the backtrader library with respect to 6 key parameters

### Entry Rules

Enter on bullish sentiment and/or technical indicators

### Exit Rules

Exit off bearish sentiment and/or technical indicators

### Risk Management

Stop losses and profit takers are not implemented due to the stability of $SPY

### Position Sizing

Default size is 100 shares, but it may also be dynamically computed. Total funds are $100k

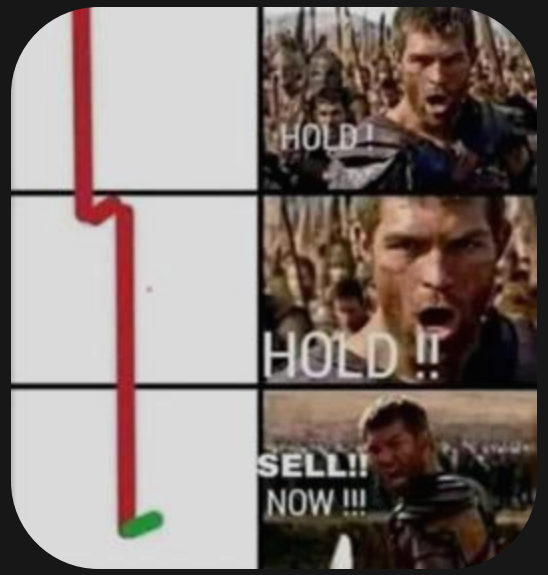
### Timeframe

The period of analysis is 11/24 to 12/19 (19 trading days)

### Market Conditions

The analysis period was done 4x to capture various market conditions

Baseline Trading Strategies




**HODL** (Hold On for Dear Life)

- **Buy and hold**


**DCA** (Dollar Cost Averaging)

- **Incremental buying** (24hrs)


**TA** (Technical Analysis)

- **Buy** if price crosses 5 EMA and RSI is oversold (<30)
- **Sell** if price crosses 10 EMA and RSI is overbought (>70)

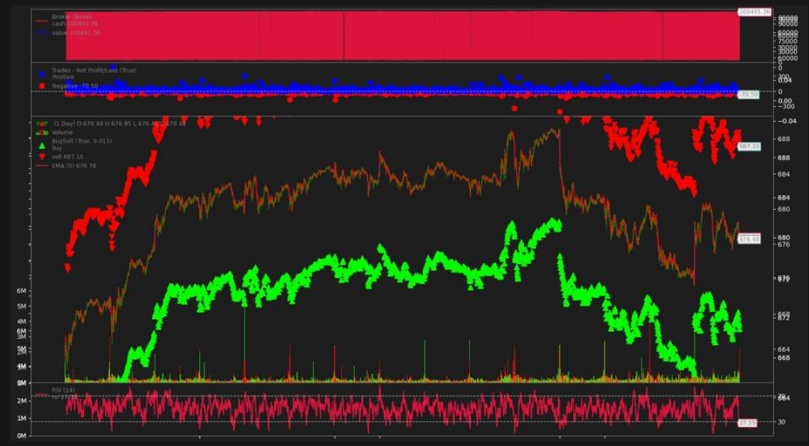
Baseline Trading Strategies – Results on the full dataset

**HODL** (Hold On for Dear Life)

- **P/L**: 2.01%
- **Sharpe**: -3.91
- **Max Drawdown**: 2.59%

**DCA** (Dollar Cost Averaging)

- **P/L**: -0.28%
- **Sharpe**: -11.12
- **Max Drawdown**: 1.11%



**TA** (Technical Analysis)

- **P/L**: 0.49%
- **Sharpe**: -6.21
- **Max Drawdown**: 1.76%

Sentiment-based Strategies

The FinBERT sentiment was used as the signal as the other methods produced subpar results

**Sentiment-only**

- **Buy**: sentiment > 0.05
- **Sell**: sentiment < -0.05

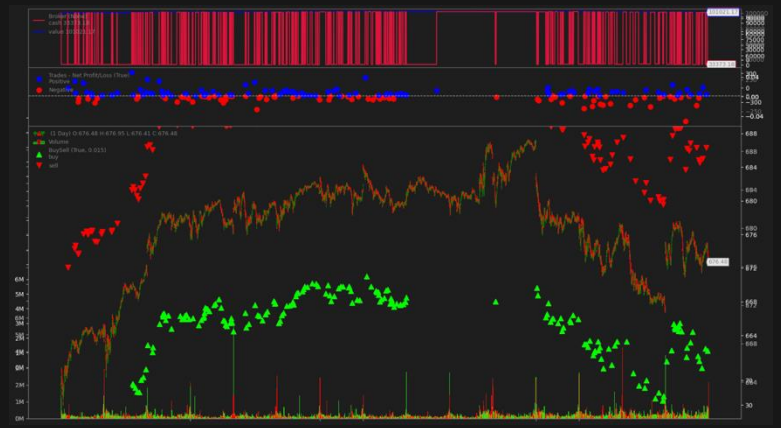
**Sentiment + TA**

- **Buy**: price > EMA and sentiment > 0
- **Sell**: RSI > 70 or sentiment < -0.1

**Inverse Sentiment**

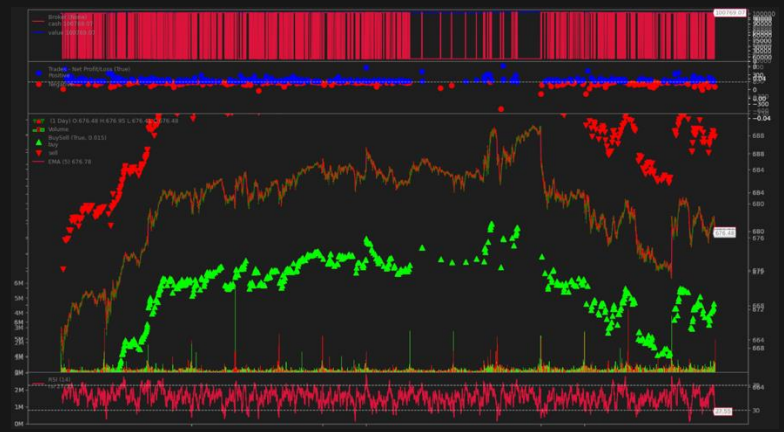
- **Buy**: sentiment < -0.15
- **Sell**: sentiment > 0.15

Sentiment-based Strategies – Results on the full dataset



**Sentiment-only**

- **P/L**: 2.06%
- **Sharpe**: -8.22
- **Max Drawdown**: 0.79%



**Sentiment + TA**

- **P/L**: 1.34%
- **Sharpe**: -8.90
- **Max Drawdown**: 0.60%

**Inverse r/wsb**

- **P/L**: -0.17%
- **Sharpe**: -8.70
- **Max Drawdown**: 1.72%

Strategy Comparison

| | Full dataset | | | Flat market | | | Bull market | | | Bear market | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P/L | Sharpe | Max Drawdown | P/L | Sharpe | Max Drawdown | P/L | Sharpe | Max Drawdown | P/L | Sharpe | Max Drawdown |
| **Buy-and-hold** | 2.01% | -3.94 | 2.59% | -0.01% | -4.23 | 1.39% | **3.79%** | -5.22 | 0.72% | **-2.18%** | -3.57 | 2.47% |
| **Incremental buying** | **-0.28%** | -11.12 | 1.11% | -0.13% | -15.89 | 0.47% | **0.43%** | -23.53 | 0.21% | -0.23% | -33.76 | 0.29% |
| **Technical analysis** | 0.49% | -6.21 | 1.76% | -0.37% | -5.78 | 1.34% | 1.79% | -8.2 | 0.73% | -1.64% | -4.74 | 1.73% |
| **Sentiment only** | **2.06%** | -8.22 | 0.79% | **1.24%** | -7.81 | 0.68% | 2.03% | -8.81 | 0.67% | **0.09%** | -12.92 | 0.75% |
| **Sentiment + TA** | 1.34% | -8.9 | 0.60% | 0.52% | -8.37 | 0.61% | 1.21% | -9.95 | 0.60% | -0.09% | -6.02 | 1.45% |
| **Inverse sentiment** | -0.17% | -8.7 | 1.72% | **-0.72%** | -11.2 | 1.16% | **0.43%** | -13.83 | 0.37% | -1.22% | -8.38 | 1.01% |

## Observations

- Sentiment-only surprisingly performed the best across ¾ of the test cases, and with relatively low drop downs
- Inverse sentiment and incremental buying performed the worst

Section 1 | Project Overview

Section 2 | Data Engineering and Scraping

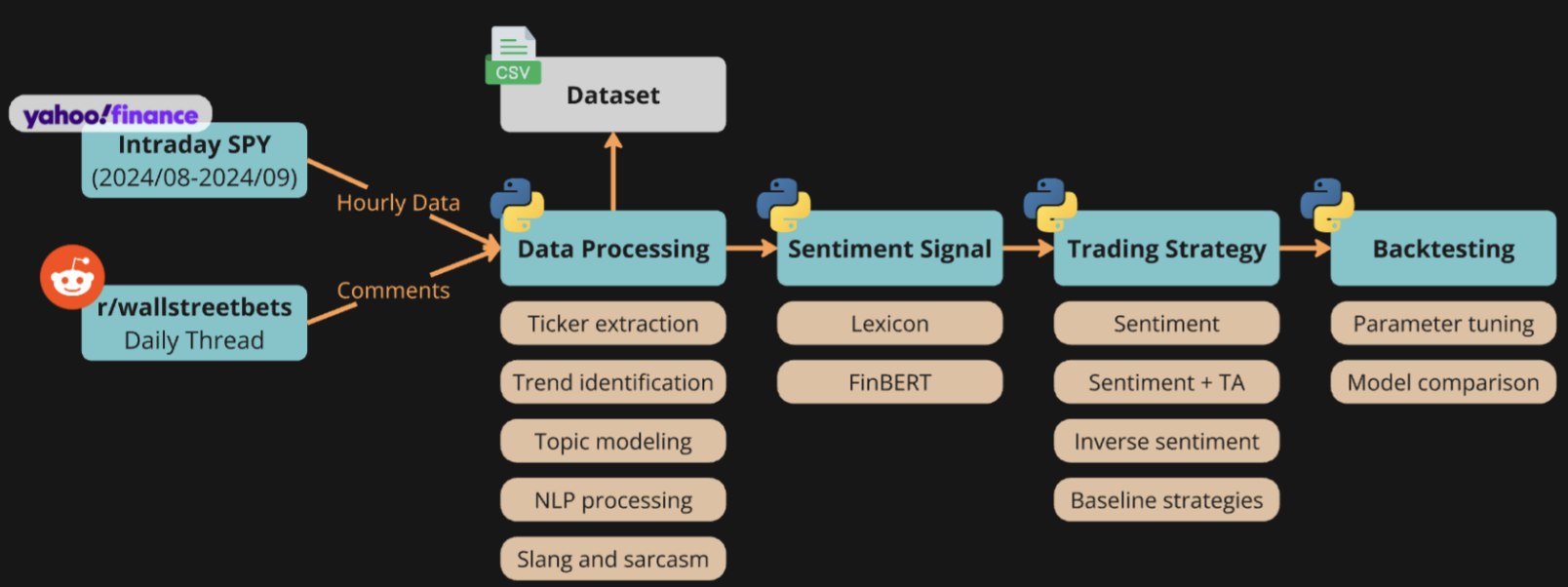
Section 3 | Sentiment Analysis

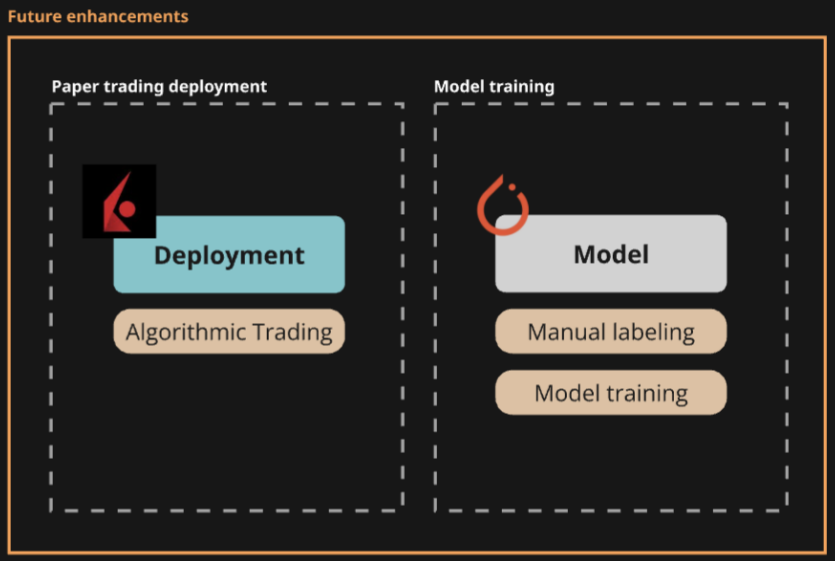
Section 4 | Trading Strategies and Backtesting

Section 5 | **Discussion**

- Limitations and Bias
- Future Enhancements
- Conclusions and Key Takeaways

Summary




## Limitations

- **Garbage in = garbage out**: The r/wsb dataset was too sparse to generate diverse signals. The data itself is very noisy
- **Unlabeled data**: Supervised sentiment analysis methods were unavailable

## Conclusions

- **r/wsb is not reliable** for sentiment-based trading due to noise, misinformation, and its reactive nature
- …however, more testing across a larger dataset is required to verify this