Predicting Stock Returns Using r/WallStreetBets Sentiment Analysis

DS 2500 – Team Project Final Report

Team members:

- Alex Zheng – zheng.alex@northeastern.edu
- Andrew Kwon - kwon.and@northeastern.edu

GitHub Repo: https://github.com/alexander-y-zheng/ds2500-final-project-reddit-sentiment-analysis

Introduction

Goal of the Project

We investigate whether social media sentiment and discussion activity on r/WallStreetBets can predict short-term stock price movements across 1-day, 3-day, and 7-day horizons.

Problem Statement & Background

r/WallStreetBets (WSB), a Reddit forum with over 15 million members, gained worldwide attention in January 2021 when coordinated retail investor activity drove massive price surges in stocks like GameStop and AMC. This event demonstrated that online communities could generate significant market momentum, challenging traditional assumptions about market efficiency. Which leads our attention to making our project examine whether WSB discussions contain predictive signals about future stock returns.

Specifically, we measure whether daily sentiment scores, mention volume, and engagement metrics correlate with subsequent price movements, and whether these metrics can forecast return direction.

Why This Matters

Individual Investors: Understand whether WSB provides signals for the stock market versus hype helps investors make decisions and avoid bubbles.

Financial Institutions: Investing firms monitor social media platforms as alternative data where our project shows whether to incorporate sentiment analysis into trading algorithms.

Dataset

Dataset Overview

Source: Reddit r/WallStreetBets (PRAW API) & Yahoo Finance(yfinance)

Time Period: September 24 - November 7, 2025 (6 weeks); extended to November 21 for forward returns

Size: 500+ posts, 72 unique tickers, 70 with price data, 1,247 ticker-day observations

Format: CSV files from API calls

Data Collection Methodology

We used PRAW to extract WSB posts, title, text, upvotes, comments, and timestamps. Stock tickers were identified via regex, and Yahoo Finance provided daily OHLC prices. Forward returns compared future closing prices to current prices across three horizons.

Variable	Description	Type	Range

Sentiment	Daily average from TextBlob	Continuous	-0.85 to +0.92
Mention count	Posts per ticker per day	Count	1 to 112
Engagement	(upvotes + 2 * comments) / mentions	Continuous	15 to 500+
Forward_return_1d/3d/7d	Future returns (%)	Continuous	-15% to +18%

Ethical Considerations

All usernames are removed immediately. WSB focuses on volatile stocks, has young male demographics and contains bots that sentiment analysis may misclassify.

Methods

Data preprocessing

We excluded posts that don't have tickers. Tickers with <3 mentions were removed. TextBlob scored sentiment (-1 to +1); daily scores were averaged per ticker. Engagement combined upvotes and 2× weighted comments, normalized by mentions. Two tickers failed to download (removed). Forward returns have missing values for the final 7 days.

Exploratory Data Analysis

WSB sentiment averages +0.08 (65% positive, 30% negative, 5% neutral) by mentions following power law: top 10% of stocks = 60% of mentions. In addition, returns show high volatility (daily $\sigma = 3.2\%$). Initial correlations: $r = 0.15$-0.22 between metrics and returns.

Modeling & Evaluation

Linear Regression predicts continuous returns using these features: sentiment, mention, engagement by evaluating with R^2 and MSE via 5-fold TimeSeriesSplit.

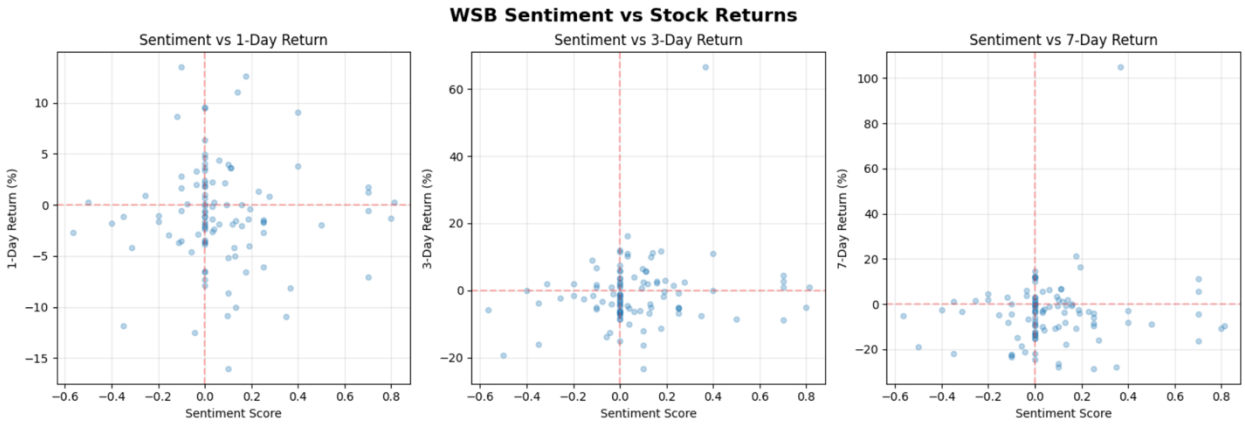
Logistic Regression predicts binary direction by evaluating accuracy vs. 50% baseline, precision, and recall.

Results and Conclusions

Finding 1: No significant correlation between WSB metrics and Returns

Sentiment for 1-day are: (return r=0.18, mention count r=0.22 and engagement r=0.19). However, the correlations declined for 3- and 7-day horizons. Even though, sentiment for 1-day shows the data this result shows that WSB captures no fundamental value which the data effects dissipate within days as fundamentals control.

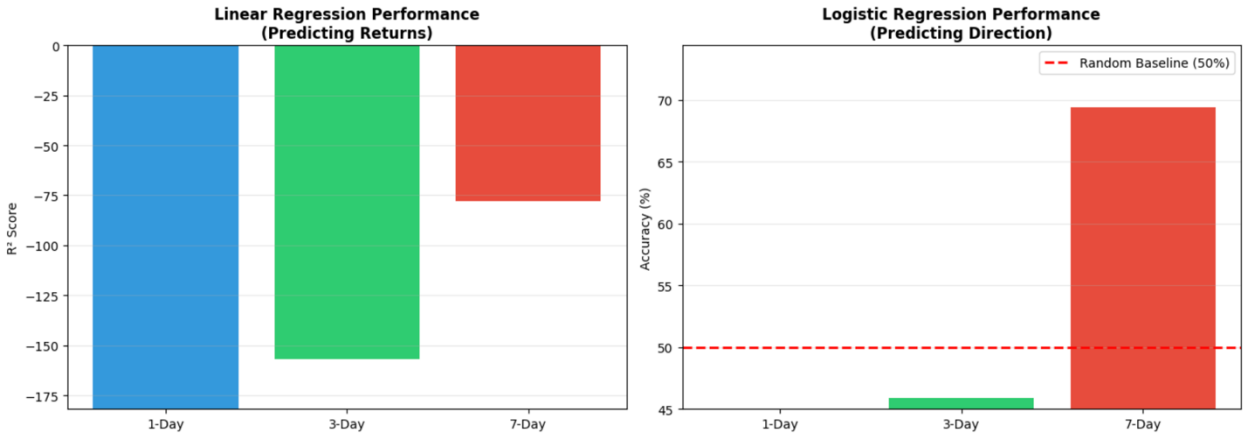
Figure1: WSB sentiment vs stock returns



Finding 2: Models fail to predict returns

Linear regression model that we produced negative R-Squared scores: -180(1 day), -155(3 day), -75(7 day), which shows that it is performing worse than predicting average return. In addition, logistic regression shows less than 45%(1 day), ~ 46%(3 day) and ~70%(7 day) which implies that the model is not working with predicting stock movement in the future.

Figure 2: negative r-squared / accuracy in different day frame



Model Performance Summary

Model	1-day	3-day	7-day	Interpretation
Linear R^2	-180	-155	-75	All perform worse than predicting mean return
Classification Accuracy	Less than 46%	~46%	~70%	Short horizons below random; 7-day likely spurious

Conclusions

Can WSB predict Stock returns? – No!

WSB sentiment showed that there is no significant correlation with returns and predicting future stock price. Model performed at or below the chance of getting it(negative R^2 score and low accuracy).

Why no signal?

1. It might because we used small sample size (99 observations)
2. It might be because of the unusual market period which in this frame the market was fluctuating because of the AI Bubble and tariffs.
3. It might be because the WSB sentiment is not working with predicting future stocks which is the best outcome that we conclude.

Implications

Retail investors should not use WSB sentiment as a trading signal. Positive sentiment coincided with negative returns, suggesting contrarian dynamics where hype peaks as prices decline.

Future Work & Limitations

Limitations

Data: Data was covered only 6 weeks which limits our tests across the market.

Select Bias: WSB focuses on volatile stocks which results in non-generalized information.

Sentiment Accuracy: TextBlob struggles with sarcasm and memes.

Future Directions

Extended Data: Longer duration data would enable regime testing.

Economic Analysis: Simulate trading strategies with transactions.

Multi-Platform: Adding Twitter or other social media will make comprehensive sentiment.

References

1. Reddit, Inc. (2025). r/WallStreetBets. https://www.reddit.com/r/wallstreetbets/
2. Yahoo Finance. (2025). Historical Stock Data. https://finance.yahoo.com/
3. Hu, D., Jones, C., Zhang, V., & Zhang, X. (2021). The Rise of Reddit. *Swedish House of Finance*.
4. Cookson, J., Lu, R., Mullins, W., & Niessner, M. (2023). Social Media and Market Fragility. *JFE*.
5. PRAW Documentation. https://praw.readthedocs.io/
6. yfinance Library. https://pypi.org/project/yfinance/