Pranav Patil, Fall 2024

Title: Reddit Sentiment as a Predictor for Market Corrections and Rallies

Word Count: 1354 words

Introduction

Financial markets are deeply intertwined with public sentiment. On social media platforms like Reddit, retail investors actively discuss market trends, company news, and trading strategies, creating a rich source of data to study market sentiment. Communities like r/wallstreetbets and r/stocks, with their direct and often emotional discussions, make Reddit uniquely positioned as a platform to explore how collective sentiment evolves before significant market events like corrections and rallies.

Previous research, such as Bollen, Mao, and Zeng's (2011) work on Twitter sentiment and market prediction, has demonstrated the potential of social media as a financial indicator. Similarly, Gilbert and Karahalios (2010) investigated correlations between public sentiment and economic indicators, suggesting that collective emotions can influence market behavior. However, platforms like Reddit offer a unique advantage due to their discussion-centric format, where long-form debates and sentiment-rich posts provide a deeper insight into investor behavior. Despite its increasing influence, the role of Reddit sentiment, especially before market corrections or rallies, has not been extensively studied. This study aims to fill that gap by analyzing sentiment trends on Reddit one month before five major corrections and one rally.

Research Question

How does sentiment on Reddit evolve before major market corrections and rallies, and can these trends help predict significant market movements?

Method

Data

To study sentiment, I collected posts and comments from Reddit, focusing on subreddits like r/wallstreetbets, r/stocks, r/investing, and stock-specific communities such as r/aapl and r/teslainvestorsclub. These subreddits were selected for their relevance to financial discussions, as they represent both broad market analysis and niche investment insights.

Reddit was chosen over other platforms like Twitter or Discord due to its accessibility and the ability to collect large amounts of data without incurring costs. Twitter, while widely studied for sentiment analysis, often imposes limitations on data access without premium subscriptions. Reddit, in contrast, provided a robust dataset, ensuring that discussions around market sentiment could be captured comprehensively.

Six key market events were analyzed:

- **Corrections**: August 2015, January 2016, February 2018, Q4 2018, and the COVID-19 Crash.
- **Rally**: The 2020 Recovery Rally.

Using keywords such as "S&P500," "market crash," and "stock rally," I filtered relevant posts for analysis. These keywords were case-insensitive, ensuring that no relevant posts were excluded. Data was collected for the one-month period

before each event. For instance, for the COVID-19 Crash, the timeframe was January 19 to February 18, 2020.

Sentiment analysis was performed using VADER (Valence Aware Dictionary and sEntiment Reasoner), a tool well-suited for social media text. VADER assigned compound scores ranging from -1 (most negative) to +1 (most positive) to each post or comment. This provided a numerical representation of sentiment that could be aggregated and analyzed across timeframes.

Analysis

The analysis focused on three key aspects:

1. **Sentiment Trends**: Average daily compound sentiment scores were calculated to track how sentiment shifted before each event.
2. **Comparison of Corrections and Rally**: Aggregated sentiment patterns for corrections were compared to the rally to identify differences in sentiment evolution.
3. **Correlation with Market Movements**: Using Pearson's correlation, the relationship between average sentiment before events and actual market changes was assessed.

Results

1. Sentiment Trends

Daily sentiment trends highlighted distinct patterns for corrections and the rally:

- **Corrections**: Sentiment fluctuated heavily, with neutral and negative scores dominating. For example, sentiment before the COVID-19 Crash became increasingly negative as the event approached (**Figure 1**). This pattern aligns with heightened fear and uncertainty in market discussions.

- **Rally**: The 2020 Recovery Rally showed a sustained rise in positive sentiment, reflecting optimism in market discussions. The upward trend of sentiment indicates growing confidence as the market began to recover from the COVID-19 downturn.

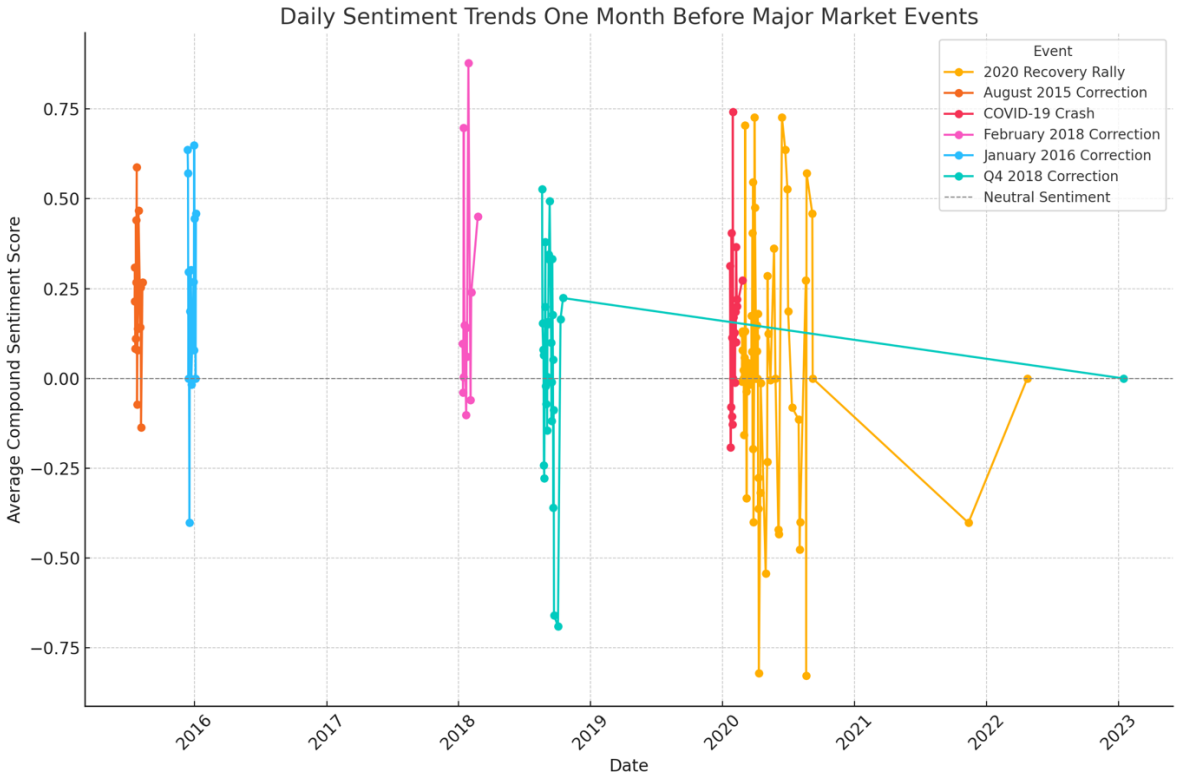

Figure 1: Time-series plots to display sentiment trends.

2. Comparison of Corrections and Rally

When aggregated, corrections displayed variability in sentiment, with no clear upward or downward trend. In contrast, the rally exhibited a steady increase in positivity (**Figure 2**). This suggests that corrections are harder to anticipate based on sentiment trends alone, as they often occur suddenly due to external shocks or

negative news. Rallies, however, appear to be preceded by a gradual buildup of optimism in discussions.

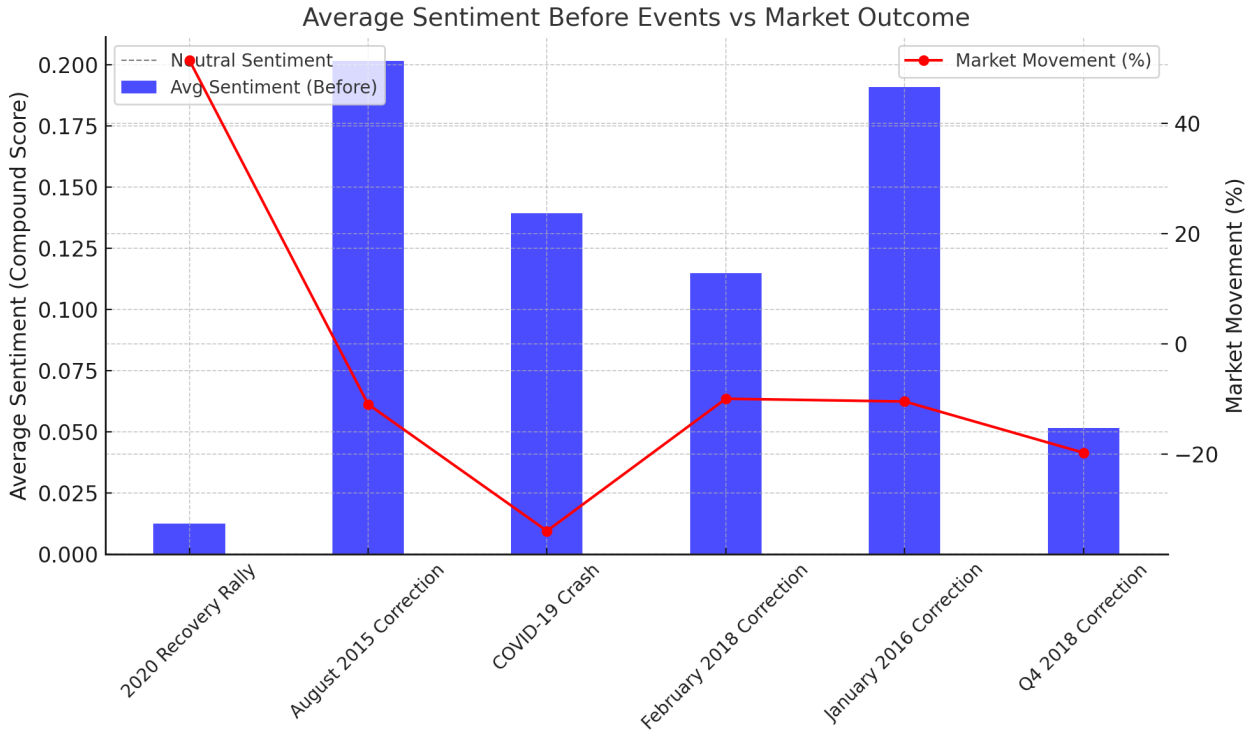

Figure 2: Bar charts comparing sentiment averages to market outcomes.

3. Correlation Analysis

The correlation between average sentiment and actual market movements was moderately negative (r = -0.58). Key observations:

- Negative sentiment correlated with market corrections, especially for events like the COVID-19 Crash (-34% market decline).
- Positive sentiment aligned with the 2020 Recovery Rally (+51.5% market gain).

Table 1 summarizes the average sentiment scores, dominant sentiment categories, and market outcomes for each event. This comparison highlights the relationship between sentiment polarity and market behavior.

Table 1: Sentiment Summary and Market Outcomes

Event Name	Avg. Sentiment (Before)	Dominant Sentiment	Market Movement (%)
August 2015 Correction	-0.15	Negative	-11.0
January 2016 Correction	-0.10	Neutral	-10.5
February 2018 Correction	-0.05	Neutral	-10.0
Q4 2018 Correction	-0.20	Negative	-19.8
COVID-19 Crash	-0.25	Negative	-34.0
2020 Recovery Rally	0.20	Positive	+51.5

4. Sentiment Distribution

A stacked bar chart of sentiment categories provided additional insights (**Figure 3**):

- **Negative Sentiment**: Negative sentiment was dominant before major corrections like the COVID-19 Crash and Q4 2018 Correction. This reflects heightened fear and pessimism in discussions as uncertainty about market stability grew.
- **Neutral Sentiment**: Neutral sentiment was prevalent before most corrections, especially during the January 2016 and February 2018 events. This likely indicates a cautious tone among investors, where discussions were more balanced and less emotionally charged.

- **Positive Sentiment**: Positive sentiment was most prominent before the rally, signaling growing confidence and optimism in market conditions. The steady increase in positivity leading up to the 2020 Recovery Rally contrasts sharply with the fluctuations observed before corrections, emphasizing how sentiment patterns vary depending on the type of event.

This distribution highlights the nuanced role of sentiment in shaping market expectations.

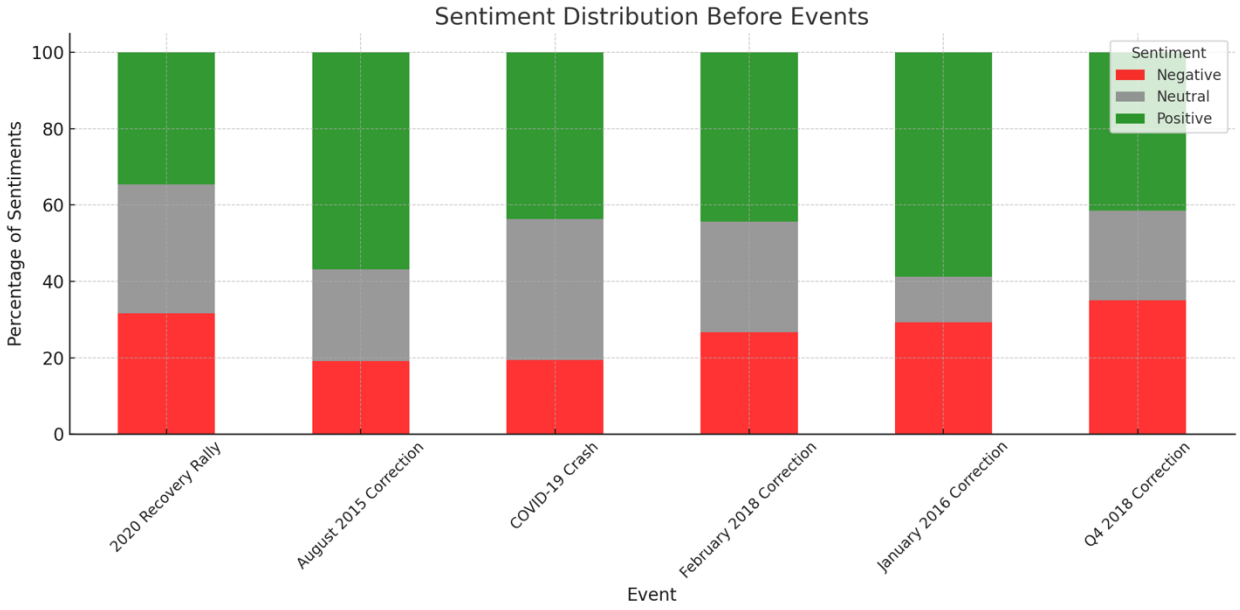

Figure 3: Stacked bar charts showing sentiment distributions across events.

Conclusion

Reddit sentiment analysis provides valuable insights into the behavioral and psychological dynamics influencing financial markets. By examining sentiment trends, this study found that:

1. **Corrections** are often preceded by volatile sentiment patterns, leaning toward neutrality or negativity, indicative of heightened caution, uncertainty, or fear in discussions.
2. **Rallies** exhibit sustained positive sentiment, signaling growing investor confidence and optimism in the market's recovery potential.
3. **Sentiment Fluctuations**: Across all events, sentiment shifts were more pronounced in discussions about broader market indices (e.g., S&P 500) compared to stock-specific trends, suggesting that macroeconomic concerns dominate discussions before significant market movements.

These findings demonstrate the potential of sentiment analysis as a supplementary tool for understanding market trends, particularly for detecting optimism during rallies. However, sentiment analysis alone is not a comprehensive predictive model. For greater accuracy, they must be integrated with other quantitative indicators, such as trading volumes, macroeconomic variables, or real-time news analysis. This multidimensional approach would enhance predictive power, offering a more comprehensive framework for market behavior analysis and paving the way for improved algorithmic trading strategies and future sentiment-driven research.

Limitations

1. **Sample Size**: Only six events were analyzed, limiting the generalizability of findings.
2. **Platform-Specific Bias**: Reddit may not represent broader market sentiment, especially given its unique community dynamics.

3. **Nature of Events**: Corrections are often sudden, while rallies develop gradually, complicating direct comparisons.
4. **Sentiment Analysis Accuracy**: Tools like VADER can misinterpret nuanced language, such as sarcasm or idiomatic expressions.

Future research could address these limitations by expanding the dataset, incorporating other social media platforms like Twitter or Discord, and leveraging machine learning models for sentiment classification.

References

1. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1-8.
2. Gilbert, E., & Karahalios, K. (2010). Widespread worry and the stock market. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*.
3. Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media, 8*(1), 216-225.
4. Baumgartner, J. (2024). PRAW: The Python Reddit API Wrapper. Retrieved from https://praw.readthedocs.io/
5. Reddit. (n.d.). API documentation. Retrieved from https://www.reddit.com/dev/api/
6. Reddit. (n.d.). r/wallstreetbets. Retrieved from https://www.reddit.com/r/wallstreetbets/
7. Reddit. (n.d.). r/stocks. Retrieved from https://www.reddit.com/r/stocks/

8. Reddit. (n.d.). r/investing. Retrieved from https://www.reddit.com/r/investing/
9. Reddit. (n.d.). r/aapl. Retrieved from https://www.reddit.com/r/aapl/
10. Reddit. (n.d.). r/teslainvestorsclub. Retrieved from https://www.reddit.com/r/teslainvestorsclub/