Reddit Sentiment as a Predictive Signal for Cryptocurrency Returns:

Evidence from Algorand

Colin Whetstone

Quant Finance Collective — Texas State University

March 2026

Abstract

We investigate whether Reddit-derived social media sentiment contains predictive information about short-term cryptocurrency price returns, using Algorand (ALGO) as our primary asset. Using 446 posts collected from r/algorand and r/CryptoCurrency over a 90-day window, we construct daily sentiment and post-volume signals and test their cross-correlation with next-day ALGO returns at lags of 0 to 7 days. We find two statistically significant correlations: post volume at lag 2 ($r = +0.24$, $p = 0.044$) and sentiment ratio at lag 3 ($r = -0.24$, $p = 0.045$). A simple timing strategy combining both signals achieves a Sharpe ratio of 1.48 and a total return of +22.21% after transaction costs over the 88-day backtest period, compared to −16.71% for buy-and-hold. A permutation test yields $p = 0.062$, providing marginal but not conclusive evidence against the null hypothesis of no predictive power. We discuss data quality limitations, look-ahead bias risks, and the conditions under which sentiment signals are likely to be most reliable.

Contents

1. Introduction

The efficient market hypothesis in its semi-strong form holds that asset prices fully reflect all publicly available information. Social media activity including posts, comments, votes, and engagement metrics represent a class of publicly available information that has grown dramatically in relevance for retail-driven markets like cryptocurrency. Unlike institutional equity markets, where information dissemination is tightly regulated, crypto markets are heavily influenced by retail sentiment, narrative, and community activity. This creates a plausible channel through which social media signals might contain predictive information about price moves.

The academic literature on social media and financial markets has grown substantially since Bollen, Mao, and Zeng [1] demonstrated that Twitter sentiment predicted Dow Jones returns several days ahead. Subsequent work has extended these findings to cryptocurrency markets, where the retail base is larger relative to institutional participation. Mai et al. [4] found that Bitcoin forum posts predicted next-day Bitcoin returns, and Kraaijeveld and De Smedt [3] documented sentiment effects across multiple cryptocurrencies.

This paper contributes to this literature by focusing on a mid-cap cryptocurrency (Algorand, ticker ALGO) rather than Bitcoin, where sentiment effects may be more pronounced due to a smaller, more concentrated community. We construct sentiment signals from two Reddit communities r/algorand and r/CryptoCurrency and test their predictive power over a 90-day window using both correlation analysis and a simple backtest with permutation testing.

2. Data and Methodology

2.1 Sentiment Data Collection

Sentiment data was collected from two Reddit communities using the public Reddit JSON API endpoint. The primary source is r/algorand, the official Algorand community subreddit, and the secondary source is r/CryptoCurrency filtered for ALGO mentions. Data collection retrieved the 100 most recent posts from each source, yielding 200 posts from r/algorand and 246 posts from r/CryptoCurrency for a total of 446 raw posts spanning approximately 90 days.

Each post was assigned a sentiment score based on its Reddit vote score: posts with positive net votes were labeled +1 (positive), posts with zero votes were labeled 0 (neutral), and posts with negative net votes were labeled -1 (negative). Raw posts were aggregated to a daily frequency, producing four daily metrics: total post count, positive post count, negative post count, and sentiment ratio (positive posts divided by total posts).

2.2 Data Quality Limitations

Several data quality limitations affect the reliability of the sentiment signal and are disclosed here for transparency.

1. Reddit vote scores skew positive by construction; heavily downvoted posts are often removed by moderators before scraping, meaning the negative sentiment category is systematically underrepresented. The median daily sentiment ratio in our dataset is 1.0 (all posts positive), which limits variation in the sentiment direction signal.
2. Reddit's search endpoint returns results sorted by recency, causing the most recent day in our dataset to contain a disproportionate share of total posts.
3. Posts may have been edited or deleted between their original publication and the time of collection, introducing a mild look-ahead bias in historical sentiment scores.
4. The 90-day window and resulting 74 days of daily observations limits the statistical power of our correlation tests.

2.3 Price Data

Price data consists of 90 days of ALGO/USDT 1-minute OHLCV data collected from KuCoin via the CCXT unified exchange library. Daily returns are computed as the percentage change in the daily closing price, resampled from 1-minute data by taking the last observation of each calendar day. The two datasets are aligned on date, retaining only days with both sentiment and price data, yielding 44 overlapping observations for correlation analysis.

2.4 Correlation Analysis

Cross-correlation between each sentiment feature and next-day returns is computed using Pearson's r at lags of 0, 1, 2, 3, 5, and 7 days. Statistical significance is assessed using a two-tailed t-test with $n - 2$ degrees of freedom. The two sentiment features tested are daily sentiment ratio (positive posts / total posts) and daily total post count, motivated by the hypothesis that post volume captures attention and interest more reliably than sentiment direction given the right-skew in our vote-based sentiment measure.

2.5 Backtest Design

A simple timing strategy is constructed using the two statistically significant findings. The strategy takes a long position in ALGO when total post count two days prior was above its trailing 14-day average *and* sentiment ratio three days prior exceeded 0.8, and holds cash otherwise. The strategy is backtested on the 90-day KuCoin price dataset with transaction costs of 0.1% per round trip applied at each trade. The statistical significance of the backtest result is assessed using a permutation test with 1,000 random shuffles of the sentiment signal.

3. Results

3.1 Cross-Correlation Analysis

Table 1 reports the cross-correlation results for sentiment ratio against next-day returns across six lag values. Table 2 reports the analogous results for total post count.

Table 1: Cross-correlation between daily sentiment ratio and next-day ALGO returns.

Lag (days)	**Correlation (r)**	**p-value**	**Significant**
0	−0.1387	0.245	
1	−0.0048	0.969	
2	−0.0165	0.892	
3	−0.2419	0.045	* ($p < 0.05$)
5	−0.0830	0.504	
7	+0.1429	0.256	

Table 2: Cross-correlation between daily total post count and next-day ALGO returns.

Lag (days)	**Correlation (r)**	**p-value**	**Significant**
0	+0.0167	0.889	
1	−0.0150	0.902	
2	+0.2412	0.044	* ($p < 0.05$)
3	+0.0025	0.984	
5	−0.0485	0.697	
7	+0.0315	0.803	

Two statistically significant correlations emerge. The sentiment ratio at lag 3 shows a negative correlation ($r = -0.24$, $p = 0.045$), indicating that high positive sentiment today is associated with slightly lower returns three days later. This is consistent with a buy-the-rumor, sell-the-news dynamic common in retail-driven crypto markets, where positive sentiment often reflects price moves that have already occurred rather than anticipating future moves. Total post count at lag 2 shows a positive correlation ($r = +0.24$, $p = 0.044$), suggesting that elevated discussion volume today predicts modestly positive returns two days later, consistent with an attention-based mechanism where increased community engagement draws new buyers into the market.

Both significant correlations have effect sizes of approximately $r = 0.24$, meaning sentiment explains roughly 6% of the variance in returns a weak but non-trivial effect.

3.2 Backtest Results

Table 3: Backtest performance summary, December 2025 – March 2026.

Metric	**Value**
Days in sample	88
Days in market	35 (39.8%)
Number of trades	14
Total return (net)	+22.21%
Buy-and-hold return	−16.71%
Outperformance	+38.92 pp
Sharpe ratio	1.484
Maximum drawdown	−17.11%
Transaction costs paid	1.40%

The strategy outperforms buy-and-hold by approximately 39 percentage points while being invested only 40% of the time. The Sharpe ratio of 1.48 is strong relative to typical passive benchmarks, though the short evaluation window and limited trade count limit the statistical interpretation of this result.

3.3 Permutation Test

The permutation test randomly shuffles the sentiment signal 1,000 times and computes the Sharpe ratio for each random shuffle. The resulting distribution has a mean of −0.524 and a standard deviation of 1.324. The actual Sharpe ratio of 1.484 exceeds 93.8% of random shuffles, yielding a permutation p-value of 0.062. This provides marginal evidence against the null hypothesis of no predictive power; the result clears the 10% significance threshold but not the conventional 5% threshold.

4. Discussion

4.1 Interpretation of Findings

The two significant correlations reflect how Reddit activity relates to ALGO price dynamics. Post volume at lag 2 suggests that community attention precedes price moves: when more people discuss ALGO on Reddit, buying activity tends to follow within 2 days. Sentiment direction at lag 3 shows the opposite sign, consistent with mean reversion after sentiment-driven price moves: when the community is uniformly positive, the move has often already been priced in, and a modest reversal follows.

These findings are consistent with the broader literature. The attention mechanism documented here is analogous to the Google Trends effects documented by Kristoufek [2] and the Twitter volume effects reported by Bollen, Mao, and Zeng [1]. The contrarian sentiment effect at lag 3 is consistent with the overreaction literature in behavioral finance.

4.2 Limitations

Several limitations constrain the conclusions that can be drawn from this analysis.

- **Sample size:** 14 trades over 88 days provides insufficient statistical power. A minimum of 50–100 independent trade observations would be needed for robust inference.
- **Vote-based sentiment:** Reddit's moderation removes heavily downvoted posts, systematically eliminating the most negative content before collection. An NLP approach using VADER or a crypto-specific model like CryptoBERT would yield more nuanced sentiment scores.
- **Look-ahead bias:** posts may have been deleted, edited, or reordered since original publication. A proper out-of-sample test would require running the signal-collection process live over several months.

4.3 Future Work

Three extensions would meaningfully strengthen this analysis:

1. Replacing vote-based sentiment with NLP-based scoring using VADER or CryptoBERT to improve measurement quality and reduce right-skew bias.
2. Extending the analysis to a 12-month window with live signal collection to increase statistical power and eliminate look-ahead bias.

3. Integrating the sentiment signal as a regime feature in the HALO Avellaneda-Stoikov market-making framework, allowing the market maker to adjust spreads preemptively when sentiment-driven volatility is detected.

5. Conclusion

We find marginal evidence that Reddit sentiment contains predictive information about short-term ALGO returns. Post volume at a two-day lag and sentiment ratio at a three-day lag both show statistically significant correlations with next-day returns ($p < 0.05$), and a simple timing strategy combining both signals outperforms buy-and-hold by 39 percentage points with a Sharpe ratio of 1.48 after transaction costs. A permutation test yields $p = 0.062$, providing evidence against the null hypothesis of no predictive power that clears the 10% but not the 5% significance threshold.

These findings are best interpreted as preliminary evidence of a weak but potentially exploitable signal rather than a definitive demonstration of predictive power. The analysis establishes the methodology and infrastructure for ongoing sentiment signal research and identifies the key improvements larger sample, NLP-based scoring, and live signal collection needed to confirm or refute the findings presented here.

References

[1] Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8.

[2] Kristoufek, L. (2013). BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports*, 3, 3415.

[3] Kraaijeveld, O. and De Smedt, J. (2020). The predictive power of public Twitter sentiment forecasting cryptocurrency returns. *Finance Research Letters*, 37, 101344.

[4] Mai, F., Shan, Z., Bai, Q., Wang, X., and Chiang, R.H.L. (2018). How does social media impact Bitcoin value? A test of the silent majority hypothesis. *Journal of Management Information Systems*, 35(1), 19–52.

[5] Shapiro, A.H., Sudhof, M., and Wilson, D.J. (2022). Measuring news sentiment. *Journal of Econometrics*, 228(2), 221–243.