Using Social Media Sentiment to Predict Bitcoin Pricing

Andrew Anantharajah
Electrical and Computer Engineering Department
Stevens Institute of Technology Hoboken, USA aanantha@stevens.edu

***Abstract*—Cryptocurrency prices exhibit extreme volatility and are strongly influenced by investor attention spans, which can be observed through social media platforms. The project aims to examine the connection between this social media sentiment, extracted from Twitter (X) and Reddit, and BTC pricing, to identify if the former can forecast the latter. The project utilizes a blend of time series model and machine learning regressors, including VADER, time series modeling via SARIMAX/ARIMA, and Gradient Boosting Regressor. Our analysis shows that the connection between sentiment score and cryptocurrency pricing is limited. It is not statistically significant over the long term, with minor gains in significance over the short term.**

I. Introduction

The frequency of cryptocurrency-related news stories and social media posts have been increasing quickly, along with the economic and societal impact of cryptocurrencies. Unlike traditional markets where price movements can be rationalized via structured financial reports and institutional decisions, cryptocurrency markets often exhibit rapid fluctuations driven by crowd sentiment, hype, and speculation. The concept of blockchain technology was initially described in [1]. The blockchain model offers a permanent record of all transactions, while offering anonymity through the use of wallet-addresses. Bitcoin's current market cap as of writing is $2.1 trillion USD. At the same time the volatility of BTC is almost 10 times that of major exchange rates [2]. In spite of the high volatility, many retail and commercial traders have BTC in their investment portfolios. In this study we look at the potential for estimating the pricing of Bitcoin using parameters that take into account past prices as well as social sentiment. While previous studies have explored technical analysis and macroeconomic factors into cryptocurrency forecasting, the extent of work on social sentiment's effects has not been fully pursued. Existing literature either focuses solely on one social media platform, uses sentiment as an explanatory variable for pricing, or applies static models without dynamic or rolling evaluations. Given the real-time and decentralized nature of cryptocurrency discussions, incorporating multi-platform sentiment analysis into predictive models is the next logical step to creating a more accurate model. Our work also extends beyond static models to test performance under rolling conditions, in order to mimic real world usage. Sentiment analysis, also known as opinion mining, is an area of study in the field of natural language processing that analyzes people's opinions, sentiments, evaluations , attitudes, and emotions via the computational treatment of subjectivity in text. For sentiment analysis we utilize the VADER model to determine the negative, neutral, positive, and compound sentiment of each social media post. We utilized this model as it is known to out-perform individual human raters when it comes to correctly identifying the sentiment of social media posts [6]. Our contributions to this problem include a unified dataset combining BTC minute level price data with Twitter and Reddit sentiment, mapped to a daily frequency. We also provide a comparative evaluation of an ARIMAX, SARIMAX, and Gradient Boosting on BTC pricing. We provide a 1-day rolling, 30-day multi-step and return-space evaluation with Diebold-Mariano tests for statistical significance of our forecast.

II. Related Works

There are several studies that explored the correlation between social media sentiment and BTC pricing. The ones we mainly looked at in the context of this paper include [3] [4] [5]. [3] has a focus on Twitter sentiment only, and using this data to train a neural network to predict pricing of BTC. [4] details the result of ARIMA modeling on Bitcoin utilizing Twitter posts as well as the Google Trends Index. [5] utilizes ARIMA modeling on various cryptocurrencies using price data and Twitter sentiment to identify which cryptocurrencies are most impacted by investor sentiment.

III. Datasets

The datasets used in this project include the Kaggle BTC historical price dataset, Reddit r/Cryptocurrency dataset, and the BTC Tweets dataset. I chose these three as they have the most wealth of data without the need for web scraping, which is not a factor in this paper. The information in these datasets that we are looking for are the Date/Time, Open, High, Low, Close, and Volume traded. We then join this data along the Data/Time with the processed Tweets and processed Reddit Data.

A. BTC Pricing Dataset

For the BTC Pricing Dataset, we had to convert the minute-level data into a daily Open, High, Low, Close, Volume dataset. We did this by aggregating the data by calculating the average values per day.

B. Social Media Datasets

For the social media datasets, we needed to clean the data to allow use of it for our models. For the BTC Tweets, the Kaggle dataset supplied two CSVs that contained tweet data, we first had to merge these two files into one, then applied our processing. The processing is as follows: We first needed to normalize all the columns to lowercase, then we filtered the text in each tweet for terms such as BTC or Bitcoin (case insensitive), then removed all punctuation, urls, and hashtags from the content, then finally we trimmed all whitespace to prevent long empty messages. We repeated the same steps for the Reddit dataset, along with removing references to external subreddits. Once the content of the messages were processed, we converted all timestamp information into Python datetime, to create a uniform time axis across the datasets. Once the datasets were fully cleaned, we ran our VADER model to create sentiment scores for each message. The scores were negative, neutral, positive, and compound sentiment scores all determined by our VADER model. We then aggregated the sentiment data for each day and combined it with our time-series dataset of BTC to create three new datasets. One for BTC Tweets with attached Twitter Sentiment, another for BTC Tweets with attached Reddit Sentiment, and another with BTC Tweets with a Compound Sentiment score. This compound score was calculated by identifying how well Twitter and Reddit sentiments correlated with BTC pricing, and gave a weight based on that. We utilized the absolute Pearson correlation between the sentiment score and BTC percentage return to determine these weights. The weight we calculated was .76 for Twitter and .24 for Reddit.

IV. Our Solution

After all the datasets were processed, we identified the accuracy of the sentiment scores compared to the price movement of BTC and found that Twitter had a sign-direction accuracy of 53.712% while Reddit had a sign-direction accuracy of 51.724%. These accuracies alone show that the social media sentiment was barely better than a coin flip at identifying the price movement of BTC. In our exploratory analysis, we modelled the range of sentiment scores, as shown in Figure. From this we can see that Twitter sentiment hovers around +0.2 while Reddit sentiment hovers near +0.5, which is an indicator of limited variance in both signals.

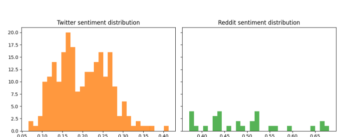

Fig. 1: Histogram of Sentiment Distribution

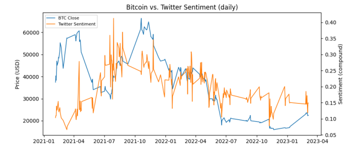

Fig. 2: Bitcoin Pricing Compared to Twitter Sentiment

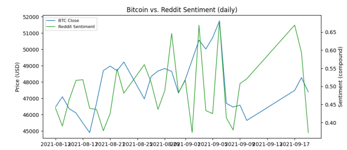

Fig. 3: Bitcoin Pricing Compared to Reddit Sentiment

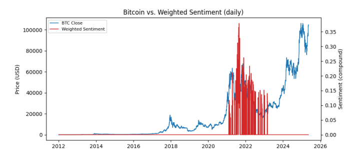

Fig. 4: Bitcoin Pricing Compared to Weighted Sentiment

A. Granger-Causality Test

We then used the Granger causality test to determine if lagged sentiment score would be able to predict the next day's BTC pricing, beyond just price analysis. This is shown in the figure.

- Let y_t denote the **BTC daily percentage return** and let x_t be a **sentiment series**.
- Maximum lag length k ($k = 7$ to span one trading week).
-

$$y_t = \alpha + \sum_{i=1}^{k} \beta_i \, y_{t-i} + \sum_{i=1}^{k} \gamma_i \, x_{t-i} + \varepsilon_t,$$

From this analysis we were able to see the following p-values for both Twitter and Reddit across different lags. We can determine from this information that Twitter sentiment had no predictive value in this dataset, and that Reddit sentiment had a small predictive sentiment at 1 day lag, which rapidly deteriorated with increased lag.

TABLE I: Granger–causality p-values $\left(H_0\text{: sentiment does } \textit{not} \text{ Granger-cause BTC returns}\right)$

Lag k (days)	p-value	
	Twitter	Reddit
1	0.9475	**0.0185**
2	0.3999	0.0612
3	0.3778	0.0505
4	0.2662	0.0846
5	0.2036	0.0749
6	0.3073	0.1405
7	0.3320	0.3070

B. ARIMAX Forecast With lag-1 Twitter Sentiment

We then used statistical analysis utilizing an ARIMA model and SARIMAX model. We chose an ARIMA model as it is a classic choice for most time-series economic models, and it is something we are more familiar with for this kind of analysis. Utilizing the preexisting price data with the bolted on sentiment data as an exogenous predictor, we found that it was a poor fit, as shown in the Figure.

TABLE II: SARIMAX(1,1,1) estimation for $\log(\text{BTC Close})$ with lag-1 sentiment

Model–fit statistics	
Number of observations	24
Log–Likelihood	47.776
AIC	-87.552
BIC	-83.374
HQIC	-86.646
Covariance type	opg

Parameter	Coef.	Std. Err.	z	p
sent_lag1	-0.1059	0.058	-1.82	0.068
ϕ_1 (AR)	0.1544	6.454	0.02	0.981
θ_1 (MA)	-0.0648	6.286	-0.01	0.992
σ^2	0.0006	0.0002	2.60	0.009

Goodness–of–fit diagnostics		
Ljung–Box $Q(1)$	0.00	$(p = 1.00)$
Jarque–Bera	8.58	$(p = 0.01)$
ARCH LM H	2.53	$(p = 0.24)$

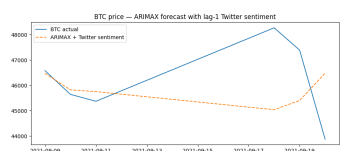

Fig. 5: Note: Does not correct for differencing resulting in malformed graph. Not important for rest of paper.

C. Best ARIMAX With lag-1 Twitter Sentiment

Then utilizing another ARIMAX model with grid search, we found that the sentiment coefficient was -.012 which carries a p-value of .906, which means it was statistically negligible.

TABLE III: Best AIC model: SARIMAX(1,0,0) for $\log(\text{BTC Close})$ with lag-1 sentiment

Model–fit statistics	
Number of observations	184
Log–Likelihood	233.910
AIC	-461.820
BIC	-452.192
HQIC	-457.917
Covariance type	opg

Parameter	Coef.	Std. Err.	z	p
sent_lag1	-0.0124	0.106	-0.12	0.906
ϕ_1 (AR)	0.9996	0.0005	2107.79	< 0.001
σ^2	0.0045	0.0002	21.14	< 0.001

Goodness–of–fit diagnostics		
Ljung–Box $Q(1)$	0.03	$(p = 0.86)$
Jarque–Bera	572.18	$(p < 0.001)$

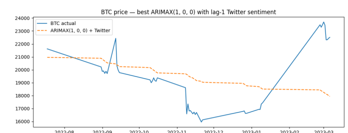

Fig. 6: MAE (log-price) = .11236

D. 30-Day Rolling SARIMAX With lag-1 Weighted Sentiment

We then ran a rolling 30-day period SARIMAX (1, 0 ,0), shown below in the Figure. This model would re-estimate each day to try and mimic real time trading. This produced a MAE value of .02617, which was worse than using just pricing data, showing sentiment scores did not aid in the forecast. Although the graph looks like it follows the pricing of BTC, this comes from the daily re-training and not from the predictive capabilities of our sentiment values.

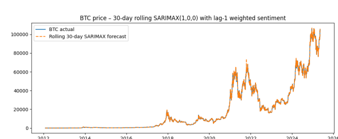

Fig. 7: MAE (log-price) = .02617

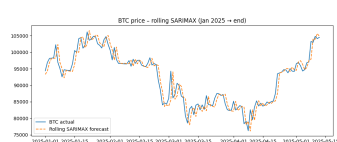

Fig. 8: Zoomed In Version

E. Gradient Boosting Regressor

The next model tested was a Gradient Boosting Regressor in order to determine the next-day log return of BTC. We used a train/test split of 80:20, with our parameters as follows: n_estimators = 200, learning_rate = .03, max_depth=3, subsample =.8, and random_state=42. The features set was as follows: [ret_lag1, ret_lag2, tw_lag1, rd_lag1]. Ret_lag1 and Ret_lag2 correspond to yesterday and the day before yesterday's returns. Tw_lag1 and Rd_lag1 are the Twitter and Reddit lagged sentiments respectively. We chose this model due to the fact that the ensemble of shallow decision trees were able to handle our nonlinear interactions between returns and sentiment. It also is able to produce feature-importance scores that we can use to identify just how much impact the sentiment scores were having on the forecast. From this we were able to see that GBR was able to create a directional accuracy of 56% which beat out both ARIMA models, however as shown in the Figure, the trend captured was very broad, and fails at following the high volatility that we expect with cryptocurrencies. Although our GBR model had the lowest MAE (MAE = .01788) compared to the other models, we see that the highest predictive power came from recent price movements. This is shown in the importance values for both ret_lag1 and ret_lag2 which accounted for 99% of the predictive power of the model.

TABLE IV: Gradient-Boosting Regressor - Feature Importance

Feature	**Importance**
`ret_lag1`	0.683657
`ret_lag2`	0.308896
`tw_lag1`	0.004557
`rd_lag1`	0.002890

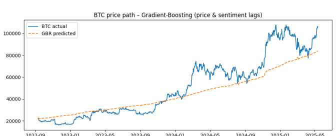

Fig. 9: MAE = .01788

F. Diebold-Mariano Test

To further prove the lack of aid sentiment scores provided, we performed a Diebold Mariano test to identify if forecast A had the same accuracy as forecast B. This was to see if any model could add value to the predictions. The baselines we used were a Lag-1 naive baseline and a zero-return baseline. The lag-1 naive baseline essentially says "Tomorrow's price will be exactly today's price," which just copies the closing value of the previous day. The zero-return baseline says "Tomorrow's return will be zero," which means the price won't go up or down on average. We compared these values to a 1-day rolling SARIMAX model, and a 30-day multi-step ARIMAX forecast. From the Diebold Mariano values, we found that the SARIMAX model augmented with sentiment data did not provide any statistical significant gains, in fact it actually degraded the one-day accuracy, while the ARIMAX augmented with sentiment data had essentially no impact with the addition of the sentiment data.

TABLE V: Error comparison and Diebold–Mariano (DM) tests

Scenario	Model	Baseline	MAE		DM	
			M	B	Stat	p
1-step	SARIMAX	Lag-1	0.0262	0.0258	2.12	0.034
30-day	SARIMAX	Const	0.0524	0.0355	—	—
Return	ARIMAX	Zero	0.0177	0.0177	1.06	0.289

V. Performance Evaluation

To summarize the performance stated above, our results show that at best, we get marginal gains from social media sentiment on pricing prediction. Raw Twitter and Reddit signals achieved only 53.7% and 51.7% directional accuracy. Our Granger-causality confirms this weakness as Twitter never provides a significant value with lags, and Reddit only provides value at a 1-day lag, that falls off immediately as the lag is increased. Our traditional time-series metrics using statistical analysis also offered no meaningful advantage with the addition of the sentiment scoring, with the 30-day SARIMAX performing worse than the lag-1 copy model, with a MAE of .02617 versus .02576. Our GBR model produced the lowest MAE of .01788 and a 59% directional accuracy, however our feature analysis shows that 98% of the predictive power came from the lagged returns, and not from the sentiment scores. Overall, our findings show that simple social sentiment alone for cryptocurrency assets is not enough to predict pricing, with the best results coming solely from the price history.

Model	Horz	Tgt	Base	MAE$_m$	MAE$_b$	p_{DM}
Logistic (lag-1 Tw)	1d	⇅	50 %	— 54 % —	50 %	—
ARIMAX Tw	1d	log P	RW	0.033	0.026	—
ARIMAX Rd	1d	log P	RW	0.031	0.026	—
Rolling SARIMAX	1d	log P	RW	0.026	0.026	0.034
30-d SARIMAX	30d	log P	Const	0.052	0.036	—
GBR (price+sent)	1d	r	0	0.024	0.027	—

TABLE VI: Model performance. RW = random-walk baseline; Horz = horizon; r = return.

VI. Our Findings

This project was aimed at determining whether daily social-media sentiment, derived using the VADER model, from Twitter and Reddit can be used to predict the future pricing of BTC. Across all models tested, from ARIMA, ARIMAX, rolling SARIMAX, and Gradient Boosting Regression, the result is that social media sentiment has little to no effect on the pricing. Twitter sentiment from our dataset was never able to become statistically significant. Reddit sentiment from our dataset is only barely statistically significant and only at a 1-day lag. The best performance we were able to get was with our GBR model, with a MAE of .01788. However, this model relied heavily on the ret_lag1 and ret_lag2 features, with very little influence from the sentiment scores. The social media sentiment scores had negligible value over price history for daily Bitcoin prediction.

VII. Future Improvements

In the future we could try to improve this project by replacing our VADER model with a different sentiment encoder, to capture more information from the social media posts. We could expand our dataset to include TikTok sentiment, although that would require video processing for sentiment analysis, which has only recently been pursued [7].

REFERENCES

[1] C. S. Wright, "Bitcoin: A peer-to-peer electronic cash system," *SSRN Electron. J.*, 2008.

[2] D. Bakas, G. Magkonis, and E. Y. Oh, "What drives volatility in Bitcoin market?," *Finance Research Letters*, vol. 50, art. 103237, 2022.

[3] H. Haritha, "Cryptocurrency price prediction using Twitter sentiment analysis," unpublished.

[4] C. Jeon, "Bitcoin and investor sentiment: statistical characteristics and predictability," unpublished.

[5] O. Kraaijeveld, "The predictive power of public Twitter sentiment for forecasting cryptocurrency prices," unpublished.

[6] C. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. Eighth Int. AAAI Conf. on Weblogs and Social Media*, Ann Arbor, MI, USA, Jun. 2014, pp. 216–225.

[7] H. Shi, "A short video sentiment analysis model based on multimodal feature fusion," *Systems and Soft Computing*, vol. 6, art. 200148, 2024, doi: 10.1016/j.sasc.2024.200148.