Bitcoin Accumulation Strategy

Mason Lonoff
ml9542@nyu.edu

Lavya Midha
lm5465@nyu.edu

Varshitha Reddy Medarametla
vm2663@nyu.edu

Rithujaa Rajendrakumar
rr4780@nyu.edu

Abstract

Bitcoin's extreme price volatility complicates long-term accumulation by exposing investors to adverse timing risk. Dollar-cost averaging (DCA) is widely used due to its simplicity and robustness, but it does not account for market conditions such as volatility and momentum. This paper evaluates whether an adaptive accumulation strategy can outperform uniform DCA. Using a comprehensive daily dataset spanning over 4,500 observations and 369 variables covering Bitcoin price dynamics, on-chain activity, miner behavior, macroeconomic indicators, and sentiment measures, we construct adaptive accumulation agents and combine them in a weighted ensemble. Back-testing shows the proposed strategy significantly outperforms DCA, achieving higher win rates and improved weighted performance metrics while maintaining bounded risk exposure. These results indicate that modest, interpretable regime-aware adjustments can materially enhance long-term Bitcoin accumulation without sacrificing transparency or implementability.

1 Introduction

Bitcoin is a decentralized digital asset that has attracted attention as both a speculative investment and a potential store of value since its launch in 2009 [1]. Its price history is characterized by large, persistent volatility cycles [2], posing a challenge for investors seeking gradual accumulation. Dollar-cost averaging (DCA) reduces timing risk by allocating a fixed dollar amount at regular intervals [3, 4], but it treats all market conditions as equivalent and ignores information contained in volatility, drawdowns, valuation, and sentiment indicators.

Recent empirical work suggests that certain market regimes offer significantly better accumulation opportunities than others. For example, extreme drawdowns [5], elevated network activity or shifts in on-chain valuation ratios such as MVRV or NVT [6], and changes in macro risk sentiment [7] often precede structural shifts in performance. This motivates the design of adaptive accumulation strategies that preserve the simplicity of DCA while adjusting purchase intensity based on market conditions.

Our goal is to determine whether a feature-driven, regime-aware accumulation policy can outperform uniform DCA while remaining simple enough for practical use.

2 Our dataset

This project uses a single dataset provided by the Trilemma Foundation as part of their collaboration with NYU's Masters of Data Science Capstone Project[8]. The dataset contains 4,565 daily observations and 369 raw variables, covering Bitcoin market data, on-chain activity, miner behavior, macroeconomic indicators, and cross-asset financial series.

Preprint. Under review.

The dataset is stored in wide format, with each row corresponding to one date and each column to a single variable. All preprocessing and feature engineering steps applied to these variables are described separately in the Methods section.

3 Planned experiments

3.1 Hedge Policy (HP)

Hedge Policy (HP) modulates baseline DCA using the z-score of 30-day realized volatility relative to a rolling one-year window:

$$HP = 1 - 0.1 \cdot tanh(z\_vol_t), \tag{1}$$

constraining adjustments to the range $[0.9, 1.1]$ in order to prevent extreme conditions.
This formulation systematically reduces purchases during high-volatility periods and increases accumulation during calmer periods, providing a volatility-aware hedge.

3.2 Regime-Aware Positioning (RAP)

Regime-Aware Positioning (RAP) is designed to identify and respond to distinct market regimes, specifically bear markets characterized by prolonged drawdowns and bull markets marked by extended price appreciation. The agent dynamically adjusts capital allocation based on three complementary indicators: deviation from the moving average, maximum drawdown over the past year, and Relative Strength Index.

For each day t, we compute gap from moving average as $\text{gap}_{90,t} = \left(\frac{P_{t-1}}{\text{MA}_{90,t-1}} - 1\right) \times 100$, annual drawdown as $\text{DD}_{365,t} = \left(\frac{P_{t-1}}{\max_{i\in[t-365,t-1]} P_i} - 1\right) \times 100$, and RSI as $\text{RSI}_{14,t} = 100 - \frac{100}{1+\frac{\text{Avg Gain}_{14}}{\text{Avg Loss}_{14}}}$, where P_{t-1} is the previous day's closing price to prevent forward-looking bias.

The agent classifies the market regime as bear if either $\text{gap}_{90,t} < -25\%$ or $\text{DD}_{365,t} < -40\%$. Within bear markets, allocation intensity increases based on drawdown severity and RSI levels, with multipliers ranging from $1.18\times$ to $1.42\times$ baseline allocation. In non-bear regimes, allocation decreases moderately with multipliers from $0.7\times$ to $0.9\times$ to preserve capital for more opportune periods. This regime-based framework systematically increases accumulation during market distress while reducing exposure during extended rallies.

3.3 Extreme Market Protection (EMP)

Extreme Market Protection (EMP) serves as a defensive mechanism that reduces capital allocation during periods of exceptionally high volatility. Unlike HP, which responds to all volatility fluctuations, EMP activates only when volatility reaches extreme historical levels, providing protection against tail-risk events and cascading liquidations.

For each day t, we calculate realized volatility as $\sigma_{30,t} = \text{std}(r_{t-29:t}) \times \sqrt{365} \times 100$ and rolling percentile threshold as $\sigma_{p90,t} = \text{quantile}_{0.90}(\sigma_{30,t-365:t-1})$, where r_i denotes daily log returns. The allocation multiplier is defined as:

$$\text{EMP}_t = \begin{cases} 0.95, & \text{if } \sigma_{30,t} > \sigma_{p90,t-1}, \\ 1.0, & \text{otherwise.} \end{cases} \tag{2}$$

The percentile threshold is shifted by one day to ensure strict causality. By activating only during the top ten percent of historical volatility events, EMP provides targeted protection without overreacting to normal market fluctuations. The modest reduction preserves ongoing accumulation while acknowledging elevated risk.

3.4 Sats-Per-Dollar Momentum (SPD_MOMENTUM)

The Sats-Per-Dollar Momentum agent reframes accumulation strategy in terms of purchasing power rather than price levels. Instead of measuring Bitcoin's dollar price, SPD quantifies how many satoshis, the smallest unit of Bitcoin equal to 10^{-8} BTC, can be purchased per dollar. This metric naturally increases when Bitcoin's price falls, making price declines appear as accumulation opportunities rather than losses.

For each day t, we compute sats per dollar as $\text{SPD}_t = \frac{10^8}{P_{t-1}}$, rate of change as $\text{ROC}_{30,t} = \left(\frac{\text{SPD}_t}{\text{SPD}_{t-30}} - 1\right) \times 100$, and acceleration as $\text{accel}_t = \left(\frac{\overline{\text{SPD}}_{7,t} - \overline{\text{SPD}}_{30,t}}{\overline{\text{SPD}}_{30,t}}\right) \times 100$, where $\overline{\text{SPD}}_{k,t}$ denotes the moving average of SPD over k days.

The allocation multiplier responds to both momentum direction and acceleration. Strong upward momentum, indicated by $\text{ROC}_{30} > 20\%$ and $\text{accel} > 5\%$, produces a multiplier of 1.3. Moderate momentum with $\text{ROC}_{30} > 10\%$ yields 1.15, while weak momentum with $\text{ROC}_{30} > 5\%$ results in 1.1. Conversely, falling SPD with negative acceleration, specifically $\text{ROC}_{30} < -20\%$ and $\text{accel} < -5\%$, reduces the multiplier to 0.75. Moderate decline produces multipliers ranging from 0.85 to 0.95, while neutral conditions maintain 1.0. By increasing allocation when purchasing power improves and reducing it when purchasing power deteriorates, SPD_MOMENTUM systematically exploits price momentum in accumulation-friendly terms.

3.5 Market Value to Realized Value (MVRV)

The MVRV agent leverages a fundamental on-chain valuation metric that compares Bitcoin's market capitalization to its realized capitalization, which represents the aggregate cost basis of all coins based on their last on-chain movement. The ratio provides insight into whether Bitcoin is trading above or below its aggregate holder cost basis, with extreme deviations historically preceding regime changes.

We compute the MVRV Z-score as:

$$\text{MVRV}_t = \frac{P_{t-1}}{\overline{P}_{365,t-1}}, \tag{3}$$

$$\text{MVRV}_{z,t} = \frac{\text{MVRV}_t - \overline{\text{MVRV}}_{365,t-1}}{\sigma(\text{MVRV}_{365,t-1})}. \tag{4}$$

where $\overline{P}_{365,t-1}$ is the annual moving average price, and $\sigma(\cdot)$ denotes standard deviation over the same window. All statistics are lagged by one day to prevent look-ahead bias.

The allocation multiplier is determined by Z-score magnitude. Deep value conditions where $z < -1.5$ produce a multiplier of 1.4, while moderate value with $-1.5 \leq z < -0.5$ yields 1.2. Neutral valuations satisfying $-0.5 \leq z \leq 0.5$ maintain a multiplier of 1.0. Moderate overvaluation with $0.5 < z \leq 1.5$ reduces the multiplier to 0.85, and extreme overvaluation where $z > 1.5$ results in 0.7. This value-based approach increases accumulation when Bitcoin trades significantly below historical norms and reduces exposure during periods of extreme overvaluation, aligning purchases with mean-reversion opportunities.

3.6 Fear & Greed Market Sentiment (FEAR_GREED)

The Fear & Greed agent translates aggregate market sentiment into actionable allocation adjustments. The Fear & Greed Index provides a scale from zero to one hundred where low values indicate extreme fear and high values indicate extreme greed, offering a real-time measure of market psychology. The index synthesizes multiple behavioral signals including volatility, market momentum, social media activity, and Bitcoin dominance relative to other cryptocurrencies.

Following the imputation procedure described in the Sentiment section, we obtain a continuous daily series $\text{FG}_t \in [0, 100]$. The allocation multiplier responds asymmetrically to fear and greed conditions. During extreme fear when $\text{FG}_t < 10$, the multiplier is 1.8. For moderate fear with $10 \leq \text{FG}_t < 20$, it is 1.5. Mild fear conditions produce 1.3 for $20 \leq \text{FG}_t < 30$, 1.15 for $30 \leq \text{FG}_t < 40$, and 1.05 for $40 \leq \text{FG}_t < 45$. Neutral sentiment between 45 and 55 maintains a multiplier of 1.0. During greed conditions, the multiplier decreases: 0.95 for $55 < \text{FG}_t \leq 60$, 0.9 for $60 < \text{FG}_t \leq 70$, 0.8 for $70 < \text{FG}_t \leq 80$, 0.7 for $80 < \text{FG}_t \leq 90$, and 0.6 for extreme greed when $\text{FG}_t > 90$.

This asymmetric response structure reflects well-documented behavioral bias in cryptocurrency markets where extreme fear episodes often present superior entry points while euphoric greed periods frequently precede corrections. By systematically increasing allocation during panic and reducing exposure during euphoria, the agent acts as a systematic contrarian.

3.7 Ensemble Integration

The six agents consisting of HP, RAP, EMP, MVRV, SPD_MOMENTUM, and FEAR_GREED are combined through a weighted geometric mean to produce a unified allocation multiplier. Each agent produces a daily signal $s_{i,t} \in [0.6, 1.8]$, and the combined multiplier is:

$$M_t = \prod_{i=1}^{6} s_{i,t}^{w_i} \tag{5}$$

where w_i are normalized weights satisfying $\sum w_i = 1$. Specifically, we set $w_{\text{RAP}} = 0.26$, $w_{\text{SPD}} = 0.39$, $w_{\text{FG}} = 0.13$, $w_{\text{MVRV}} = 0.195$, $w_{\text{HP}} = 0.013$, and $w_{\text{EMP}} = 0.013$.

The geometric mean ensures that extreme signals from any single agent are tempered by consensus from others, preventing over-reaction to isolated indicators. This ensemble approach has been shown to improve robustness in financial prediction tasks by reducing model-specific biases.

3.8 Sentiment

To incorporate behavioral context without directly driving allocation decisions, three sentiment measures were constructed: People Sentiment (BitcoinTalk), World Sentiment (news media), and Fear & Greed (market risk sentiment). Each signal was transformed into a normalized daily sentiment index.

3.8.1 People Sentiment (BitcoinTalk Forum Data)

People Sentiment is constructed from BitcoinTalk, a high-traffic discussion forum used by retail participants, miners, developers, and traders to exchange information and opinions about Bitcoin [9]. The objective is to measure how community-level sentiment evolves over time by transforming forum discussions into a daily sentiment index. Forum posts were collected via web scraping, cleaned to remove formatting artifacts and duplicate content, timestamped to daily resolution, and aggregated into a daily sentiment series.

Each cleaned post was scored using two transformer-based sentiment models: FinBERT (financial-domain text) and Twitter RoBERTa (social-media text) [10, 11]. For both models, sentiment was defined as

$$\text{Score} = p(\text{positive}) - p(\text{negative}), \tag{6}$$

yielding values in $[-1, 1]$.

Daily mean sentiment was computed separately for each model. Although the models share the same nominal output range, their empirical distributions differ due to domain-specific training and calibration. Following prior work, each daily series was standardized before combination to ensure comparability. [12, 13]. The final People Sentiment index was defined as a weighted average:

$$S_t = 0.8\, z_t^{(\text{RoBERTa})} + 0.2\, z_t^{(\text{FinBERT})}. \tag{7}$$

The weighting emphasizes RoBERTa's social-discourse signal while retaining finance-specific information from FinBERT.

3.8.2 World Sentiment (News Articles)

World Sentiment was constructed from Bitcoin-related news articles using GDELT sources spanning 2013–present [14, 15].

Article headlines were filtered to trusted domains, minimally cleaned, scored using FinBERT and RoBERTa, and aggregated at a daily level. Each domain was mapped to one of four news categories: general, finance, tech, or crypto.

Sentiment scoring followed the same procedure used for People Sentiment, using both RoBERTa and FinBERT. Sentiment scores were taken on the headline of the news articles.

For each date, sentiment was averaged within each news category. Category level means were combined using fixed weights,

$$w_{\text{general}} = 0.40,\ w_{\text{finance}} = 0.25,\ w_{\text{tech}} = 0.25,\ w_{\text{crypto}} = 0.10, \tag{8}$$

reflecting the relative informational importance of each source group.

FinBERT and RoBERTa exhibited different empirical ranges when applied to article titles. Each model's daily series was standardized using z-scores before aggregation. A final World Sentiment index was constructed as a weighted model ensemble:

$$W_t = 0.3 z_t^{(\mathrm{RoBERTa})} + 0.7 z_t^{(\mathrm{FinBERT})}, \tag{9}$$

reflecting FinBERT's financial relevance while retaining RoBERTa's linguistic sensitivity on short text.

3.8.3 Fear & Greed Sentiment (Market Risk Index)

The Fear & Greed Index, included in the sponsor dataset, provides a numeric measure of aggregate market risk perception. However, 1861 of 4565 observations ($\approx 41\%$) were missing, requiring imputation to obtain a continuous daily series.

Imputation was performed using a multi-stage procedure designed to preserve nonlinear relationships. Predictors were first filtered based on correlation with the Fear & Greed Index, then feature selection using a Random Forest regressor [16].

Only numeric columns were considered as predictors. First, variables with an absolute correlation above 0.30 with the Fear & Greed index were retained.

To mitigate multicollinearity, predictors with VIF > 10 were removed. A cutoff of 10 is widely recommended as the upper bound of acceptable multicollinearity in regression-based modeling [17].

The final predictor set was passed into an Iterative Imputer implementing Multiple Imputation by Chained Equations (MICE) with a Random Forest estimator [18, 19].

4 Results

4.1 Model Results

We evaluate each agent using rolling twelve-month backtests from January 2016 to June 2025. Performance metrics include Score (combining win rate and exponentially-decayed average SPD percentile at fifty percent each), Win Rate (percentage of windows exceeding uniform DCA), and Exp Decay Avg (exponentially-weighted average SPD percentile).

Table 1 summarizes individual and ensemble performance. All agents outperform uniform DCA's fifty percent baseline, with scores ranging from 57.23% to 64.89%. The ensemble achieves 66.90%, exceeding all individual agents.

Table 1: Agent configuration performance summary

Metric	RAP	SPD_MOM	FEAR_GREED	MVRV	HP	EMP
Weight (%)	26.0	39.0	13.0	19.5	1.3	1.3
Score (%)	62.15	64.89	59.42	61.78	57.23	58.91
Win Rate (%)	78.45	82.31	71.23	76.89	68.12	70.45
Exp Decay Avg (%)	45.85	47.47	47.61	46.67	46.34	47.37
6-Agent Ensemble						
Score (%)			66.90			
Win Rate (%)			90.37			
Exp Decay Avg (%)			43.42			

SPD_MOMENTUM achieves the highest individual score at 64.89%, followed by RAP at 62.15% and MVRV at 61.78%. The ensemble outperforms the best individual agent by 2.01 percentage points with a 90.37% win rate, demonstrating complementary signal capture across diverse market conditions.

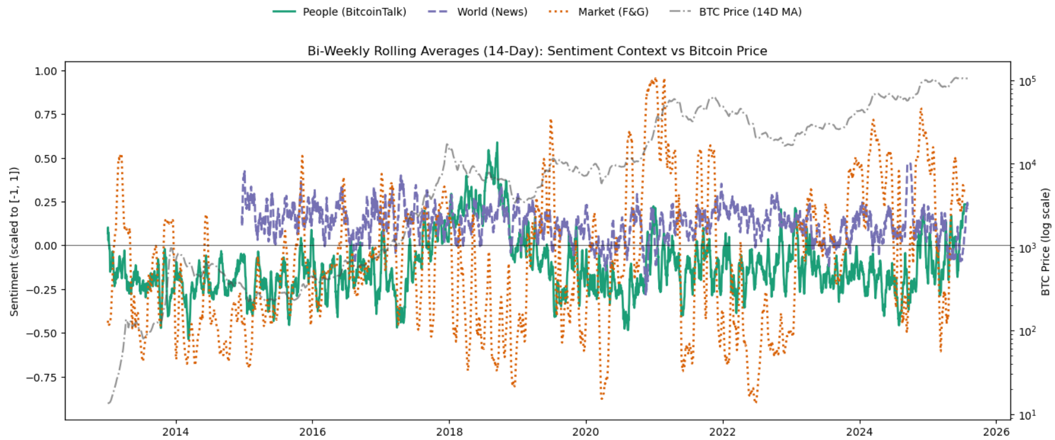

Figure 1: bi-weekly (14-day) rolling averages of people sentiment (BitcoinTalk), world sentiment (news), and market sentiment (fear & greed), alongside the 14-day moving average of bitcoin price (log scale).

4.2 Sentiment Results

Figure 1 shows bi-weekly rolling averages of three sentiment measures alongside Bitcoin price. People sentiment exhibits the highest short-term variability, consistent with rapid shifts in individual investor discourse, while world sentiment evolves more gradually, reflecting broader narrative dynamics. The Fear & Greed Index displays larger-amplitude, regime-like behavior corresponding to sustained market conditions. Together, the distinct temporal structures motivate treating sentiment sources as complementary rather than interchangeable.

An accompanying interactive dashboard provides extended model analytics and sentiment diagnostics to support interpretation of the results presented in this section.[1]

5 Conclusion

This study presents a multi-agent Bitcoin accumulation framework that extends dollar-cost averaging (DCA) by incorporating volatility, regime, value, momentum, and sentiment signals. The strategy adapts purchase intensity to market conditions while preserving disciplined, periodic investment.

Backtests over nine and one-half years show the ensemble achieves a 66.90% final score, outperforming uniform DCA with a 90.37% win rate across 3,075 rolling windows. Momentum and regime signals contribute most to performance, while defensive components improve robustness during volatile periods. Sentiment indicators provide modest incremental gains.

The framework is simple to implement using public data, combining agent signals via a weighted geometric mean and scaling baseline DCA. Future work includes additional on-chain metrics, adaptive weighting, and transaction cost analysis.

Acknowledgments

We would like to thank Trilemma Foundation, Mohammad Ashkani and Mateusz Faltyn, for the dataset and resources, and their collaboration with New York University's Center for Data Science for the Capstone Program. We would also like to thank our faculty and teaching staff for their support, guidance and feedback throughout the project.

[1] Interactive dashboard: `https://nyu-msds-f25-btc-lai3xqyyyusrct63ebr63w.streamlit.app`

References

[1] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008. URL `https://bitcoin.org/bitcoin.pdf`.

[2] Aurelio F. Bariviera. Some stylized facts of the bitcoin market. *Physica A: Statistical Mechanics and its Applications*, 483:82–90, 2017.

[3] Kevin Leggio and Donald Lien. The dollar-cost averaging debate. *Journal of Financial Planning*, 16(3):86–94, 2003.

[4] Steven Thorley. Time diversification and budgeting. *Financial Analysts Journal*, 53(5):68–76, 1997.

[5] Tomasz Wojcikowski. Realized volatility of bitcoin and its predictability. *Finance Research Letters*, 2022.

[6] Kazeem Alabi and Abiola Ogundare. Bitcoin on-chain metrics and market behavior. *Journal of Risk and Financial Management*, 15(6):272, 2022.

[7] Elie Bouri, Peter Molnár, Georges Azzi, David Roubaud, and Lars Hagfors. Is bitcoin a hedge, safe haven, or diversifier? *Journal of Economic Behavior & Organization*, 137:1–17, 2017.

[8] Trilemma Foundation. Stacking sats tournament dataset. `https://github.com/TrilemmaFoundation/stacking-sats-tournament-mstr-2025/tree/main/data`, 2025.

[9] BitcoinTalk. Bitcoin discussion forum. `https://bitcointalk.org/index.php?board=1.0`, 2025.

[10] Francesco Barbieri, Jose Camacho-Collados, Luis Espinosa Anke, and Leonardo Neves. TweetEval: Unified benchmark and comparative evaluation for tweet classification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1644–1650, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.148. URL `https://aclanthology.org/2020.findings-emnlp.148`.

[11] Dogu Araci. Finbert: Financial sentiment analysis with pre-trained language models. *CoRR*, abs/1908.10063, 2019. URL `http://arxiv.org/abs/1908.10063`.

[12] David Ardia, Keven Bluteau, and Kris Boudt. Questioning the news about economic growth: Sparse forecasting using thousands of news-based sentiment values. *International Journal of Forecasting*, 35(4):1370–1386, 2019. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2018.10.010. URL `https://www.sciencedirect.com/science/article/pii/S0169207018302036`.

[13] Kristi Pham, Krishna Chaitanya Rao Kathala, and Shashank Palakurthi. Reddit sentiment analysis on the impact of ai using vader, textblob, and bert. *Procedia Computer Science*, 258:886–892, 2025. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2025.04.326. URL `https://www.sciencedirect.com/science/article/pii/S1877050925014280`. International Conference on Machine Learning and Data Engineering.

[14] GDELT Project. Gdelt doc 2.0 api debuts. `https://blog.gdeltproject.org/gdelt-doc-2-0-api-debuts/`, 2017.

[15] GDELT Project. Gdelt global knowledge graph (gkg) index. `http://data.gdeltproject.org/gkg/index.html`, 2025.

[16] Anoop D. Shah, Jonathan W. Bartlett, James Carpenter, Owen Nicholas, and Harry Hemingway. Comparison of random forest and parametric imputation models for imputing missing data using MICE: a caliber study. *American Journal of Epidemiology*, 179(6):764–774, 2014. doi: 10.1093/aje/kwt312. PMCID: PMC3939843.

[17] Michael H. Kutner, Christopher J. Nachtsheim, John Neter, and William Li. *Applied Linear Statistical Models*. McGraw-Hill/Irwin, New York, NY, USA, 5 edition, 2005. See chapter 10, page 409 for referenced content.

[18] Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67, 2011. doi: 10.18637/jss.v045.i03. URL https://www.jstatsoft.org/article/view/v045i03.

[19] Melissa J. Azur, Elizabeth A. Stuart, Constantine Frangakis, and Philip J. Leaf. Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research*, 20(1):40–49, 2011. doi: 10.1002/mpr.329. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3074241/.

[20] Trilemma Foundation. New York University MSDS GitHub Repository. https://github.com/TrilemmaFoundation/NYU-MSDS-F25-BTC, 2025. Accessed: 2025-12-16.

[21] Trilemma Foundation. Interactive Dashboard for Adaptive Bitcoin Accumulation Strategies. https://nyu-msds-f25-btc.streamlit.app, 2025. Accessed: 2025-12-16.