Can On-Chain Analysis Predict Bitcoin and Cardano Returns?

Methodology for testing predictive claims in cryptocurrency markets

Gabriel Mazzali Garcia
Undergraduate Research
XP Educação, Bachelor of Data Science
Advisor: Professor Pedro Calais

July 2026

Project repository:

https://github.com/GabrielMazzali/onchain-vs-price

Contents

Abstract

This project asks a narrow question: after accounting for information already visible in market prices, do on-chain metrics improve forecasts of future Bitcoin and Cardano returns? The study uses daily data from January 2020 to June 2026. It compares **price-based** variables with network activity, valuation ratios, miner data, exchange flows, sentiment, and staking information (for Cardano).

The analysis was divided into two steps. The first tested whether the data contained a relationship with future returns. The second tested whether statistical and machine learning models could use any relationship that remained. This separation matters because a weak model and a weak signal can produce the same poor result. The analysis also examined volatility, long-horizon cycle timing, price momentum, and market regimes.

For the daily prediction tasks tested here, the answer was consistent. On-chain variables did not provide a reliable improvement over price alone for either asset. Some results looked positive before statistical corrections, but they disappeared after accounting for overlapping return windows, repeated testing, time trends, leakage, and simple baselines. On-chain data was still useful for describing market conditions, and a long-horizon Bitcoin valuation result was promising, but the available history contained too few independent market cycles to treat it as established evidence.

The main contribution of this project is not a trading rule. It is a clear example of how financial prediction claims should be evaluated, so that a **good-looking** result is not mistaken for real forecasting value.

The central comparison

Model A used only information derived from price.

Model B used the same dates, the same validation windows, and the same model, but added on-chain variables.

If Model B did not improve consistently, on-chain data did not add useful predictive information beyond price.

1. Introduction

1.1 Why On-chain

Public blockchains record transactions, supply changes, network activity, and other information that does not exist in the same form in traditional markets. This has led to a popular claim among cryptocurrency investors and analysts: **since on-chain metrics show what investors and network participants are doing, they should reveal future price movements before the market price does.**

The claim is plausible, but plausibility is not enough. Many dashboards describe the current market very well. A high MVRV ratio, for example, can indicate that holders have large unrealized profits. That does not automatically mean the next return will be negative. Description and prediction are different tasks, and this project was designed to keep them separate.

A negative result is useful only when the test gives the signal a fair chance to appear. For that reason, the analysis did not begin with the most complex model available. It first examined if a relationship existed in the data, then if models could use it, and finally if the answer changed across prediction horizons or model types.

1.2 Questions raised

The two questions that started the project:

Question	What it means in practice
RQ1. Is there a raw signal in the data?	Do current on-chain values have a measurable relationship with future returns before a model is fitted?
RQ2. Can a model use it?	Do on-chain variables improve predictions when compared with the same model using price alone?

These questions matter because a model that fails is ambiguous: “there is no signal” and “the model is too weak” look identical from the outside. Answering RQ1 independently of any model helps separate these two explanations. As the investigation progressed, other questions were added:

Question	What it means in practice
RQ3. Can the metrics predict move size?	Even if direction is hard, can on- chain data improve forecasts of future volatility?
RQ4. Does it help with cycle timing?	Do valuation and sentiment indicators work better over many months than over days?
RQ5. What does work?	Do price momentum or market regime descriptions contain useful information?
RQ6. Does using other techniques change the answer?	Can an LSTM find a sequence pattern missed by simpler models?

1.3 The answer in one paragraph

Using the selected metrics and daily prediction horizons, the study found that on-chain data did not improve forecasts beyond the information already see**n** in price. This result was supported by both machine learning models and model-free statistical tests, reducing the chance that it was caused by a limitation of one specific method. The analysis also identified some useful findings: past price showed a small momentum effect, on-chain variables helped identify different historical market regimes, and Bitcoin presented a promising long-term valuation pattern during one major cycle. These results suggest that on-chain data may be more useful for understanding market conditions and long-term behaviour than for reliably predicting daily returns.

2. Background and Related Work

2.1 What the main metrics represent

On-chain metrics summarize different parts of a blockchain network. The table below gives the practical meaning of the variables most important to this study.

Metric	Meaning	Typical interpretation
MVRV	Compares the current market value with the value of coins at the prices where they last moved.	A rough measure of unrealized profit and long-term valuation.
NVT	Compares network value with transaction activity.	A valuation ratio similar in spirit to comparing a company value with business activity.
Puell Multiple	Compares current miner revenue with its one-year average.	Can indicate unusually high miner profits or periods of miner stress.
Hash rate	Measures the computing power used to secure Bitcoin.	Shows network investment and miner participation.
Active addresses and transaction intensity	Measure how much the network is being used.	Describe network activity and participation.
Exchange flows	Measure coins moving into or out of exchanges.	Often interpreted as possible selling or buying pressure.
Fear and Greed Index	Combines market indicators into a sentiment score.	Describes whether market behavior is fearful or optimistic.
Cardano staking ratio	Measures the share of ADA participating in staking.	Describes network participation and the amount of supply committed to staking.

2.2 Why previous studies often report very high accuracy

Many cryptocurrency forecasting studies report excellent fit, especially when using LSTM networks or other deep learning models [6]. A common reason is that the model predicts the future price level rather than the future return. These are not equivalent tasks.

Bitcoin prices are highly persistent from one day to the next. A model can therefore predict that tomorrow will be close to today and receive an R^2 value near 1.0. For example, one recent hybrid study reported an R^2 of 0.9941 for Bitcoin price prediction [7]. The result looks impressive, but it may contain no information about whether the price will rise or fall. A useful test must compare the model with the simple rule “tomorrow equals today” and must also evaluate returns, not only price levels. Section 5.7 reproduces this exact problem with the LSTM used in this project.

2.3 The efficient-market theory

There is also a theoretical reason to expect a negative result. On-chain data is **public.** Under even a weak form of market efficiency [13], information that is freely available and genuinely predictive tends to be arbitraged into the price. A metric like MVRV can therefore honestly *describe* the market’s current state (“richly valued relative to cost basis”) without *predicting* its next move. This alone does not prove the claim to be negative but explains why extraordinary claims (R^2 = 0.99) demand critical examination. A more accessible explanation of this theory and its practical implications is provided by Coelho [18].

3. Data and Features

3.1 Data sources and period

The analysis uses daily Bitcoin and Cardano data from January 1, 2020 to June 1, 2026, giving 2,344 observations for each asset. Market and blockchain variables came from the Coin Metrics community API [1]. Cardano staking data came from Blockfrost [2] and was converted from epoch observations to daily values. Market sentiment came from the Crypto Fear and Greed Index [3]. All feature engineering was placed in one shared module to ensure that every model and statistical test used the same definitions, dates, missing value rules, and lag construction.

The code used for data preparation, modelling, and evaluation is available in the project's GitHub repository.

3.2 Feature groups

Group	Examples	Role in the experiment
Price variables	Log return, distance from 7, 30, 90, and 182-day averages, and 30-day volatility	Forms the baseline that on-chain data must beat.
Network activity	Active addresses, transaction intensity, and money velocity	Measures current use of the blockchain.
Valuation	MVRV and NVT	Measures market value relative to holder cost or network activity.
Mining and supply	Hash rate, issuance in US dollars, Puell Multiple, and fees	Describes miner economics and network supply.
Flows and sentiment	Exchange inflow, outflow, net flow, and Fear and Greed	Describes exchange movement and market mood.
Cardano-specific	Circulating supply, active stake, and staking ratio	Adds variables specific to the Cardano network.
Total	37 variables for Bitcoin and 30 for Cardano	Includes 20 lagged versions of four base series.

Some variables called on-chain are partly built from price. MVRV and NVT include market capitalization, and issuance measured in US dollars multiplies coin issuance by price. Fear and Greed is a market sentiment indicator rather than a pure blockchain measure. These variables were kept because they are widely used, but the price-only baseline is necessary to prevent their price component from being mistaken for a separate on-chain signal.

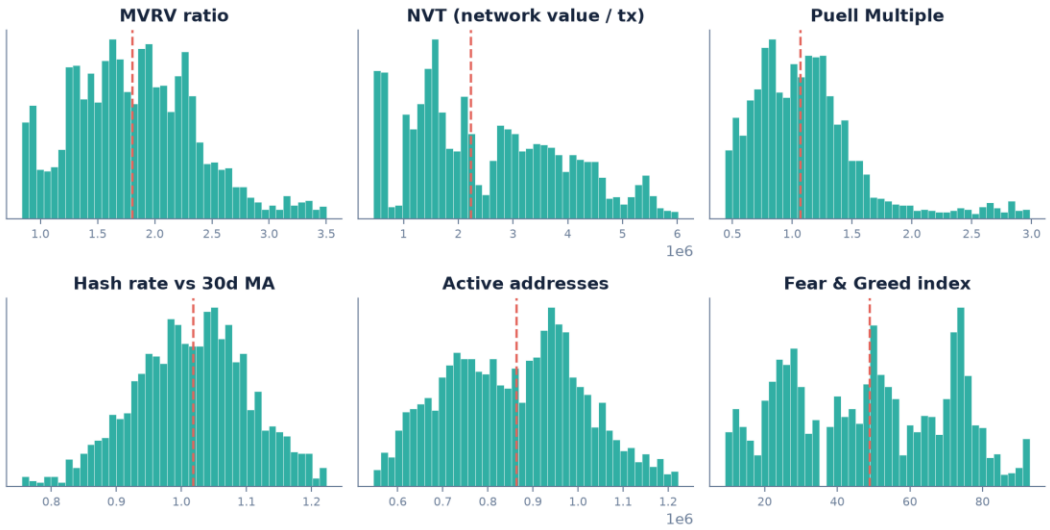

Figure 1. Distributions of six representative Bitcoin metrics. The dashed line shows the median. The shapes differ greatly, which is one reason a single modelling assumption may not suit every variable.

3.3 Lags and leads

A model needs information from the recent past, so four base series were copied at lags of 1, 3, 7, 14, and 30-days. A 30-day lag, for example, gives the model the value that was known 30-days earlier. The future targets were returns over 3, 7, 14, and 30 days. Rows at the end of the data set were removed when their future return was not yet known. They were never assigned an artificial label.

For classification, future returns were converted into Buy, Hold, or Sell categories using predefined thresholds. The deep learning comparison used the Bitcoin 30-day target with a 1 percent threshold. The exact label choice is later discussed as a limitation because the Hold class became small.

Figure 2. A base metric (active addresses, 7-day smoothed) and its 30-day lag: the same shape shifted right. Lagged features let a model reference the network's recent past; forward returns are the leads it tries to predict.

4. Methodology

The methodology is the heart of the project. Financial time series break the assumptions of ordinary machine learning, and each safeguard below blocks a specific problem. Together they are what make the negative result trustworthy.

4.1 Walk-forward validation and anti-leakage rules

Random train and test splits are inappropriate for this task because they allow the model to learn from the future and then be evaluated on the past. Instead, every model used walk-forward validation. The model trained on an early period, was tested on the next period, then moved forward and repeated the process. This imitates how the model would have been used in real time.

1. **The scaler was fitted separately within each training window.** The mean and standard deviation of the future test period were never used to transform the training data.
2. **Unknown future labels remained unknown.** Rows without a complete future return were dropped rather than converted into a valid class.
3. **Model selection also respected time order.** Hyperparameter search used time-series splits instead of shuffled folds.

4.2 How it was measured

The report includes the **F1 score** for each class and uses the **Matthews Correlation Coefficient, or MCC**, as the main summary score [10]. MCC ranges from negative 1 to positive 1. A value near zero means that the

predictions are not meaningfully better than chance. It is useful here because class proportions are uneven and a model should not receive a high score simply for predicting the most common class.

Scores were reported for every walk-forward fold, not only as one final average. A model with an average score of 0.15 and a standard deviation of 0.15 is not producing a stable advantage. The analysis also tracked errors where Buy was predicted as Sell or Sell as Buy because those mistakes are more costly than confusing either class with Hold.

4.3 The core experiment: price alone versus price plus on-chain data

The key test was a controlled ablation. For each configuration, the same model was trained twice on the same rows and validation folds. The first version used only price variables. The second added the **on-chain** variables. The difference between the two scores represents the contribution of the added data. A Wilcoxon signed-rank test was then used to check if these differences were consistently positive across the validation periods, rather than being caused by a few isolated results.

Why this comparison is necessary

An on-chain model can appear successful because several on-chain ratios already contain price.

It can also succeed because price itself is predictable at a very weak level.

Only a matched price baseline shows whether the blockchain variables add information that was not already available.

4.4 Model-free tests: Is the signal even there?

Three tests were used to determine whether a relationship existed before fitting a predictive model. Each has a characteristic blind spot discussed in Section 7:

- **Lead-lag correlation.** This measured whether a current feature moved with a future return. It is simple and easy to interpret, but it can miss nonlinear relationships and can give misleading p-values when future windows overlap.
- **Granger causality.** This tested if a feature's past values improved a forecast of the next return *beyond* the return's own history [8].
- **Mutual information.** This looked for any statistical dependence between a feature and the direction of future returns, including nonlinear relationships. The observed values were compared with results obtained after shuffling the target labels.

4.5 Two corrections that decide everything

The first correction concerns overlapping windows. A **30-day** future return calculated every day shares 29 days with the next observation. Those observations are not independent, so an ordinary p-value treats the sample as much larger than it really is. This issue was corrected with Newey West standard errors [9].

The second correction concerns repeated testing. When many features, horizons, assets, and models are tested, some results will cross the usual 5% significance threshold by chance. Bonferroni and Benjamini-Hochberg false discovery rate corrections were therefore applied to the collection of results [17, 11].

4.6 Models used

The supervised models were logistic regression and XGBoost [14]. The deep learning stage added an LSTM sequence classifier [4] and a simple DLinear-style baseline [5]. Both neural models were implemented in PyTorch [16] and trained through Poutyne [15]. Class imbalance was handled with inverse-frequency weights

instead of oversampling, since synthetic observations could distort the time structure, and duplicating minority-class examples could make the models overfit to a few historical periods.

5. Results

5.1 RQ1: Was there a signal in the data? (No.)

The three model-free tests agreed after their specific weaknesses were addressed.

- **Lead and lag correlations looked stronger before correction**. The naive analysis found 67 significant correlations for Bitcoin and 77 for Cardano. After correcting for overlapping future returns, 28 and 23 remained. They were weak, with absolute correlations below 0.34, and were concentrated in valuation ratios that already include price.
- **No on-chain variable passed the Granger test.** The only feature that consistently survived was Bitcoin's 7-day price momentum measure.
- **Mutual information selected variables that acted like a calendar.** Circulating supply and the moving average of hash rate were strongly correlated with time, so they could identify when an observation occurred rather than predict what happened next.

5.2 RQ2: Could the supervised models use the data? (No.)

Logistic regression and XGBoost produced an MCC of about 0.09. More importantly, the variation from one validation fold to another was about as large as the average score. A model that works in one period and fails in the next does not provide reliable evidence of prediction.

The ablation tested 24 combinations of assets, horizons, and models. In no configuration was the improvement from on-chain data larger than its own fold-to-fold variation. On average, the measured improvement was about six times smaller than the noise around it. One of the 24 comparisons had an uncorrected p-value below 0.05. With 24 tests, about 1.2 such results are expected by chance. That single result did not survive either multiple testing correction.

The small effect also changed sign across models and configurations. Sometimes the added variables helped slightly and sometimes they hurt slightly. The simplest interpretation is that the difference was random variation, not a repeatable advantage.

5.3 RQ3: Did on-chain data improve volatility forecasts? (Still no.)

Volatility tends to cluster, so future move size can be more predictable than return direction. The models did capture part of that persistence. However, none of the 12 comparisons between price-only and price-plus-on-chain models remained significant after the overlap correction.

At the 30-day horizon, the model initially appeared to reduce error by 22% to 34%. However, after accounting for the strong overlap between consecutive 30-day periods, the result was no longer statistically significant, with p-values of approximately 0.11 to 0.12. Further analysis showed that most of the apparent gain came from trending variables rather than blockchain behavior.

5.4 RQ4: Did the indicators work better for long-term cycle timing?

This question was added later in the project because on-chain indicators can also sometimes be used for long-term cycle analysis. At horizons up to 180 days, the valuation and sentiment rule underperformed buy and hold. For Bitcoin, it produced a CAGR of −2.8%, compared with +41.5% for buy and hold.

Over a 540-day horizon, the pattern was more promising. Periods with low MVRV and high fear produced total returns of 137%, compared with 91% for the unconditional baseline, while expensive and greedy periods produced 44%.

However, the result was driven mainly by one episode around the 2022 market bottom and did not repeat for Cardano. It should therefore be treated as a possible long-term pattern, not as a confirmed finding.

5.5 RQ5: What did show useful information?

Price momentum was real but economically weak

The 7-day price momentum variable was the only feature to pass the Granger test, and only for Bitcoin. However, as a simple trading rule, it did not produce strong results. The strategy held Bitcoin only when its price was above the moving average. Its win rate per trade was between 18% and 31%, while its daily direction was correct only about half the time. The 7-day rule returned 28% per year, compared with 43% for buy and hold, even before fees.

Longer moving-average rules gave smoother returns relative to their risk, with a Sharpe ratio of 1.31 versus 0.90 for buy and hold. However, the best window was chosen after checking several options on the same data, so it may only look good in hindsight. In practice, momentum helped more by reducing large losses than by correctly predicting whether the price would rise or fall next.

On-chain variables described market regimes

Unsupervised clustering grouped the historical data into recognizable states such as accumulation, healthy activity, euphoria, and distribution. Those groups had different average forward returns. At 30-days, the Bitcoin groups ranged from positive 5.5 percent to negative 0.9 percent, while the Cardano groups ranged from positive 36.9 percent to negative 12.8 percent.

This is useful for explaining what kind of market was present in the historical sample. It is not yet a prediction result because the groups and their returns were evaluated on the same data used to create them. To claim forecasting value, the cluster definitions need to be fixed on past data and then tested on a later period.

5.6 RQ6: Did the LSTM change the conclusion?

Does a modern sequence model succeed where linear and tree models failed? An LSTM classifier and a DLinear-style linear baseline were tested. The results are summarised:

Model	Balanced F1	MCC	Opp.-dir. errors
LSTM (sequence)	0.36	+0.17	34%
DLinear (linear floor)	0.18	+0.06	14%
XGBoost (reference)	0.41	+0.09	—

The LSTM had a higher MCC than XGBoost, but it did not have a higher balanced F1 score. Its Buy class F1 was 0.51 with a standard deviation of 0.34, showing that performance varied greatly across validation periods. Its opposite-direction error rate was 34%, meaning that it often confused Buy with Sell or Sell with Buy. The model leaned toward the common Buy class and paid for that choice with costly mistakes.

DLinear provided a useful lower reference but performed poorly. The overall conclusion is that additional model capacity moved the result from poor to weak and unstable, not from weak to reliable. A sequence model cannot create information that is absent from the inputs.

5.7 The persistence illusion

The regression version of the LSTM demonstrates why a price-forecasting study can report an excellent score and still have no useful predictive skill. When asked to predict the future price level, the model achieved an R^2 of 0.98, visually a near-perfect fit (Figure 3, left). When the same model was evaluated on returns, its R^2 was −0.74, with 52% directional accuracy, which is close to chance, and it performed worse than the naive "tomorrow equals today" baseline (Figure 3, right).

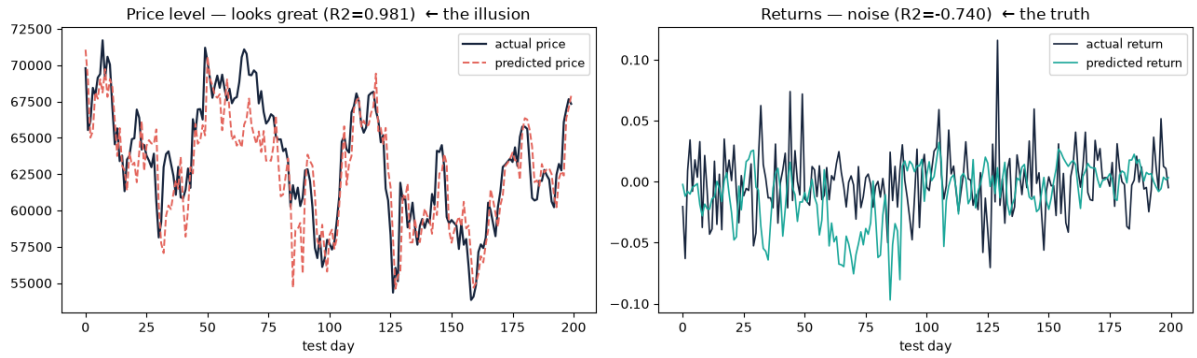

Figure 3. Predicting the price level looks accurate because tomorrow is usually close to today. Predicting the return reveals that the same model has no useful advantage.

The important lesson is simple. A high R-squared score on a persistent price level does not show that a model can predict returns or trading direction. The result must be compared with a persistence baseline and evaluated on a quantity that represents the decision being made.

6. Why no daily predictive advantage was found

The experiments do not prove that no on-chain variable can ever predict any cryptocurrency at any horizon. They show that the tested variables did not add a detectable advantage for daily Bitcoin and Cardano prediction during this period. Three explanations fit the evidence.

1. **Efficient markets.** On-chain data is public. Predictive public information tends to be arbitraged into price [13]. A metric can describe the state of the network without forecasting its next move.
2. **Several variables already contain price.** MVRV, NVT, and issuance in US dollars move partly because price moves. They can appear predictive until they are compared with a price-only baseline.
3. **Daily direction may be the wrong horizon for some indicators.** The 540-day Bitcoin result was more consistent with the intended use of valuation metrics than the daily tests. The current sample simply did not contain enough independent cycles to evaluate that claim with confidence.

7. Six common traps in financial prediction studies (and how they were addressed)

This section directly answers the question "why do other studies say otherwise?" Several positive-looking results were reproduced and then corrected in this project. Other traps, such as global scaling and random splitting, were prevented by the study design rather than tested as separate experiments. Together, the six traps form a checklist for evaluating predictive claims in financial time series.

#	The trap	What it faked here	The fix
1	**Overlapping windows**	Lead-lag flagged 67–77 "hits"; volatility p = 1e−18	**Newey-West p-values**

#	The trap	What it faked here	The fix
2	**Multiple testing**	1 of 24 “wins” by luck	**Bonferroni and FDR**
3	**Trend / clock effect**	Supply and hash rate acted like hidden date variables, creating apparent MI and volatility gains	**Check correlation with time**
4	**No price baseline**	A model could look successful because some on-chain variables already contained price	**Price-only ablation**
5	**Leakage**	Global scaling, random split, synthetic labels	**Per-fold scaling and walk-forward validation**
6	**Persistence illusion**	LSTM: $R^2 = 0.98$ by learning that tomorrow is close to today	**Predict returns; beat a naive baseline**

These checks are not specific to cryptocurrency. They are useful whenever a study uses time series, many variables, many model configurations, or a persistent target. A result should remain convincing after these simpler explanations have been removed.

8. Limitations

The conclusions should be read within the boundaries of the available data and design.

- **One-and-a-half cycles (2020–2026).** The long-horizon Bitcoin result is therefore a case study around one main bear market bottom, not a test across several independent cycles.
- **ADA staking data requires a leakage check.** Epoch-level values were converted to daily data, so it must be confirmed that a value was not assigned to days before it became available.
- **The 1 percent threshold used in the 30-day classification made Hold observations rare, at about 6 percent.** A triple-barrier label based on stop-loss, take-profit, and time limits would be a useful robustness test [12].
- **The descriptive regime result is in-sample.** It was not evaluated as a fixed clustering rule on a later period.
- **The scope includes two assets, daily data, and one historical period.** The methods can be reused elsewhere, but the numerical conclusions should not automatically be generalized to every asset, frequency, or indicator.

9. Conclusions and Future Work

9.1 Conclusions

For Bitcoin and Cardano at the daily horizons tested from 2020 to 2026, on-chain metrics did not provide a detectable predictive improvement over price itself. Direct statistical tests did not find a clear raw signal. Logistic regression, XGBoost, and the LSTM did not obtain a stable advantage when the variables were added to a price baseline. The volatility, momentum, and economic backtests led to the same practical conclusion.

This is not the same as saying that on-chain data has no value. It described historical market regimes, provided context about valuation and network conditions, and produced a promising long-horizon Bitcoin pattern. The

distinction is that these uses were descriptive or not supported by enough independent cycles to establish a forecast.

The most important result is methodological. Several analyses initially appeared successful. They stopped looking successful when the tests accounted for overlap, repeated comparisons, time trends, leakage, price baselines, and persistence. A credible forecasting claim should survive all those checks.

Final answer

Past price contained a small amount of predictive information, but not enough to create a strategy after costs.
The tested on-chain variables did not add a reliable daily forecasting advantage beyond price.
Their clearest value in this study was explaining the current market state, not predicting the next return.

9.2 Future work

- Extend the Bitcoin history to approximately 2011 so that the long-horizon valuation test includes several major cycle bottoms.
- Repeat the classification with triple-barrier labels to test whether the result depends on the current Buy, Hold, and Sell definition [12].
- Use text-based sentiment, where language models are better matched to the data, and compare it with the existing numeric sentiment index.
- Complete the look-ahead audit of the Cardano staking pipeline before drawing final conclusions about staking variables.
- Fix the regime definitions using an early period and evaluate them on a later period to separate description from forecasting.

References

[1] Coin Metrics. (n.d.). Coin Metrics API v4: Community data. https://docs.coinmetrics.io/api/v4

[2] Blockfrost. (n.d.). Cardano blockchain API. https://blockfrost.io

[3] Alternative.me. (n.d.). Crypto Fear & Greed Index. https://alternative.me/crypto/fear-and-greed-index/

[4] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[5] Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023). Are Transformers Effective for Time Series Forecasting? Proceedings of the AAAI Conference on Artificial Intelligence, 37(9), 11121–11128. https://doi.org/10.1609/aaai.v37i9.26317

[6] Ahmad, S., Shah, S., & Mehmood, G. (2026). A Systematic Literature Review of LSTM Networks for Bitcoin Price Forecasting Performance. Spectrum of Engineering Sciences, 4(6), 210–223.

[7] Daş, B., Dagdogen, H. A., Kaya, M. O., & Das, R. (2026). A new hybrid approach combining transformer-based language models and graph neural networks for cryptocurrency forecasting. Engineering Applications of Artificial Intelligence, 164, 113179. https://doi.org/10.1016/j.engappai.2025.113179

[8] Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica, 37(3), 424–438.

[9] Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica, 55(3), 703–708.

[10] Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta, 405(2), 442–451.

[11] Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B, 57(1), 289–300.

[12] López de Prado, M. (2018). Advances in Financial Machine Learning. Wiley.

[13] Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25(2), 383–417.

[14] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

[15] Paradis, F., Beauchemin, D., Godbout, M., Alain, M., Garneau, N., Otte, S., Tremblay, A., Bélanger, M.-A., & Laviolette, F. (2020). Poutyne: A Simplified Framework for Deep Learning. https://poutyne.org

[16] Paszke, A., et al. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. Advances in Neural Information Processing Systems, 32, 8024–8035.

[17] Dunn, O. J. (1961). Multiple Comparisons among Means. Journal of the American Statistical Association, 56(293), 52–64. https://doi.org/10.1080/01621459.1961.10482090

[18] Coelho, S. (2026). O que é a Hipótese dos Mercados Eficientes? Edward Knings, Substack. https://open.substack.com/pub/edwardknings/p/o-que-e-a-hipotese-dos-mercados-eficientes?r=fo2je&utm_medium=ios