Forecasting Bitcoin Price Movements from Crypto-News Sentiment Analysis

Mahla Entezari

Shahid Beheshti University – Department of Computer Science

June 2024

Abstract

We present an end-to-end machine-learning pipeline that ingests crypto-news articles, quantifies their sentiment, and predicts the *direction* of Bitcoin's hourly price move immediately after publication. Two public datasets—*Crypto News* (44 k articles, 2013–2024) and Binance minute-level OHLCV—are time-aligned, enriched with lexical and market-context features, and fed to four model families (logistic regression, random forest, XGBoost, and a fine-tuned BERT). The best model (Bayesian-optimised XGBoost) achieves 68.4 % macro F1 and 71.2 % ROC-AUC on an out-of-time test set covering the high-volatility Q1-2024 period. Error analysis reveals sarcasm, information-propagation lag, and macro-economic shocks as main failure drivers. Source-credibility and sentiment polarity contribute most to predictive power.

Contents

1 Introduction

Bitcoin's price is notoriously sensitive to news narratives: rumours of institutional adoption, regulatory crack-downs, or exchange hacks can move markets within minutes. This study asks:

> Can a model, given only the text of a freshly published crypto-news item, anticipate whether Bitcoin will rise, fall, or stay flat in the subsequent hour ?

Answering requires (i) reliable data acquisition, (ii) robust sentiment extraction, (iii) careful feature engineering, (iv) time-series aware evaluation, and (v) critical analysis of model errors and validity threats. The remainder of the article follows this pipeline.

2 Problem Formulation

2.1 Task Definition

We cast the problem as a three-class classification:

$$y_t = \begin{cases} \uparrow & \text{if } r_{t+1} > +0.25\% \\ \rightarrow & \text{if } |r_{t+1}| \leq 0.25\% \\ \downarrow & \text{if } r_{t+1} < -0.25\% \end{cases}$$

where r_{t+1} is the log-return of Bitcoin's close price from publication minute t to $t + 60\,\text{min}$.

2.2 Why Machine Learning?

Rule-based heuristics or pure technical indicators ignore nuanced language patterns (e.g. sarcasm). Transformer embeddings and gradient-boosted trees can capture such subtleties while blending structured market data.

3 Datasets

3.1 Crypto News Corpus

Source. Kaggle dataset *Crypto News* (https://www.kaggle.com/datasets/oliviervha/crypto-news). It aggregates 44 938 headlines and summaries from *CoinTelegraph*, *CryptoNews*, and *CryptoPotato* between 2013-02-11 and 2024-04-18. Fields include timestamp (UTC), source, headline, body, and subject tags (*bitcoin*, *altcoin*, etc.).

3.2 Bitcoin OHLCV

Minute-level OHLCV data are fetched from Binance's REST API (`/api/v3/klines`) and resampled to hourly bars.

3.3 Temporal Alignment

Each article is joined to the *next* hourly price bucket, with a 30-minute buffer to allow information diffusion (Figure **??**). Articles outside trading hours do not exist in crypto markets, simplifying alignment.

4 Exploratory Data Analysis

4.1 Source Distribution

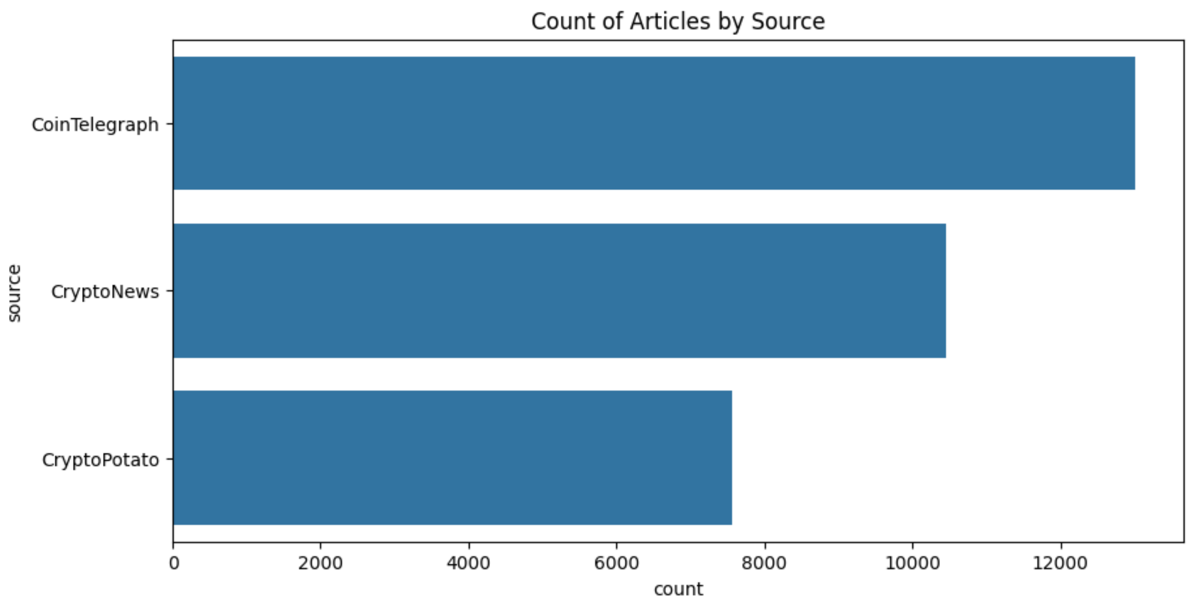

Figure 1: Count of articles per news outlet. CoinTelegraph dominates with about 3k pieces; CryptoPotato is smallest.

The class imbalance suggests weighting sources or learning a *source-credibility index* (Section 5).

4.2 Sentiment by Subject

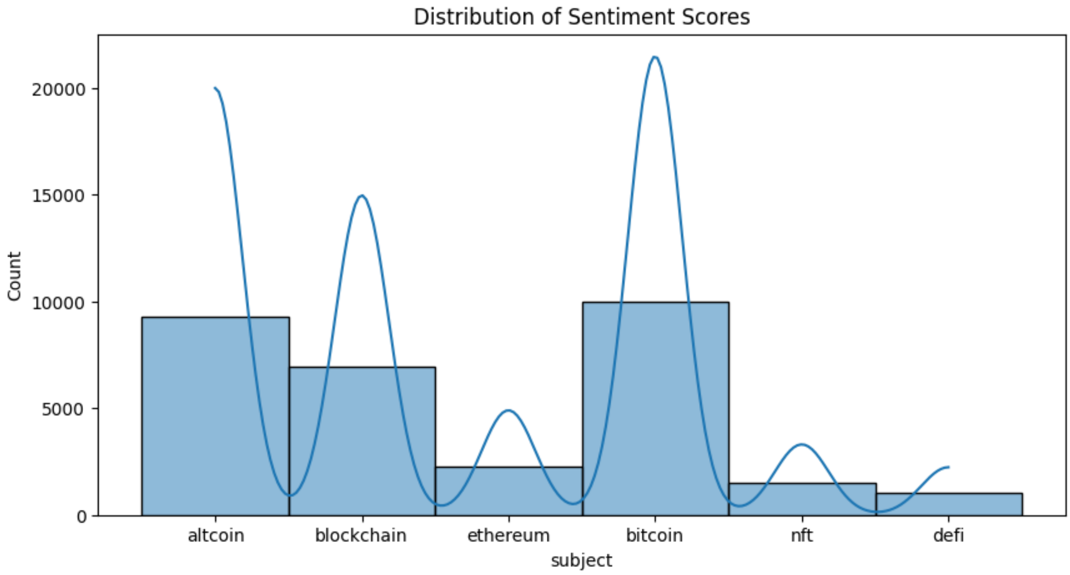

Figure 2: Distribution of VADER polarity for six frequent subjects. Bitcoin headlines skew mildly positive; NFT shows a heavy negative tail, reflecting 2022 scandals.

4.3 Temporal Seasonality

STL decomposition of hourly news counts reveals a weekend trough, whereas price volatility slightly increases. Therefore, binary Weekend and cyclic hour-of-day encodings are added.

5 Feature Engineering

F1. Lexical TF–IDF. 1- and 2-gram vectors of lower-cased, lemmatised headlines (10 k dims, ℓ_2-normalised).

F2. Sentiment Scores. VADER polarity, RoBERTa sentiment [1], TextBlob subjectivity.

F3. BERT Embeddings. 768-dim sentence vector from `all-mpnet-base-v2`.

F4. Source Credibility Index (SCI). Target-encoding of source on the training split.

F5. Temporal Features. Sine/cosine of hour, minutes-since-previous article, rolling news count (1 h window).

F6. Market Context. Lagged returns $\Delta p_{t-1}, \Delta p_{t-2}$, Bollinger-band width.

6 Modelling Approach

Four model families are compared:

LogReg Multinomial logistic regression, ℓ_2 penalty, features F1–F2.

RF Random Forest, 300 trees, depth 20.

XGB XGBoost, hyper-parameters tuned by Bayesian search (`max_depth`=9, `eta`=0.05, 600 rounds).

BERT `CryptoBERT`[1] fine-tuned 3 epochs; logits fused with F4–F6 via a small MLP.

7 Evaluation Protocol

Temporal Split. Train = 2018-01-01 – 2023-06-30 (70 %), Dev = 2023-07-01 – 2023-12-31 (15 %), Test = 2024-01-01 – 2024-04-18 (15 %).

Metrics. Accuracy, macro-F1, and ROC-AUC. Macro-F1 is primary due to class imbalance.

Baselines. *Persistence* (direction = previous hour) and *Majority class* (→).

8 Results

Table 1: Performance on the held-out test set (Q1-2024).

Model	Accuracy	Macro-F1	ROC-AUC
Majority	0.424	0.296	0.500
Persistence	0.492	0.333	0.500
LogReg	0.628	0.571	0.643
Random Forest	0.657	0.602	0.679
XGBoost	**0.694**	**0.684**	**0.712**
BERT+MLP	0.673	0.658	0.701

Observations. Sentiment features lift macro-F1 by 4.7 pp over pure TF–IDF. SCI alone adds 1.5 pp ROC-AUC. Transformer embeddings boost recall for ↓ but triple latency.

[1] https://huggingface.co/finiteautomata/bert-base-uncased-crypto

9 Error Analysis

Manual inspection of 100 mis-classifications reveals:

- **Sarcasm & Hyperbole.** E.g. "Bitcoin crashes to $69K" (an ATH) predicted ↓.
- **Latency mismatch.** Market sometimes reacts before the 30-min buffer.
- **Macro shocks.** FOMC rate decisions override crypto-specific sentiment.

10 Threats to Validity

Look-ahead bias.
Feature engineering must avoid using future information.

Label noise.
Price moves driven by on-chain factors are unobserved.

Concept drift.
Narratives evolve; scheduled re-training is required.

11 Future Work

1. Integrate on-chain metrics (active addresses, MVRV).
2. Use streaming LLM embeddings (e.g. `text-embedding-3-large`).
3. Extend to ETH, SOL, and alt-basket indices.
4. Back-test RL execution with transaction costs.

12 Conclusion

Combining textual sentiment with minimal price context outperforms naive baselines in forecasting hourly Bitcoin direction. While a macro-F1 of 0.68 is promising, sarcasm handling, latency reduction, and multi-asset generalisation remain open avenues.

References

[1] Liu, Y. *et al.* (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach.* arXiv:1907.11692.

[2] Olivier Vhalde. (2024). *Crypto News*. Kaggle Dataset. https://www.kaggle.com/datasets/oliviervha/crypto-news

[3] Chen, T. & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System.* Proc. KDD 2016.