Name: Manan Amit Shah

Cryptocurrency Price Direction Prediction

Executive Summary

Overview

This project develops a machine learning model to predict next-day price direction for ten major cryptocurrencies using technical indicators and logistic regression. The model processes 30 days of recent historical data with 21 engineered features, trained using expanding window validation to prevent look-ahead bias.

Goals

The research addressed three core objectives:

1. Can standard indicators (RSI, MACD, moving averages, Bollinger Bands) predict daily cryptocurrency direction?
2. Does a single approach work across diverse assets with different volatility profiles?
3. How do we validate time series models without using future information?

Methodology

Data: 60 days of OHLCV data per cryptocurrency from Yahoo Finance. Extra data ensures indicator warm-up periods (14+ days for RSI, 20+ days for Bollinger Bands).

Features: 21 technical indicators across four categories:

- Momentum: RSI-7, RSI-14, MACD, ROC
- Trend: SMA-7, EMA-7, SMA-14, EMA-14, moving average crossovers
- Volatility: 7-day rolling std, Bollinger Band width and position
- Interactions: RSI × MACD, momentum/volatility ratio, lagged features

Targets: Two binary labels—next-day direction (up/down) and significant moves (±0.5%-0.8% thresholds).

Model: Logistic regression with L2 regularization and balanced class weights. Chosen for interpretability, computational efficiency, and robustness.

Validation: Expanding window approach where training set grows and test set always predicts one day ahead, preventing look-ahead bias.

Key Results

- **Average F1-Score:** 0.37
- **Average Precision:** 0.32
- **Average Recall:** 0.43
- **Best Performers:** Ethereum (F1: 0.44), Litecoin (F1: 0.44), XRP (F1: 0.40)
- **Significant Moves:** F1-scores reached 0.50-0.55 for Litecoin and XRP

Results demonstrate modest predictive signal above random guessing, but insufficient for reliable standalone trading.

Key Limitations

Data Constraints:

- 30-day window insufficient for different market regimes
- No macroeconomic, on-chain, or sentiment data
- Technical indicators capture only ~20% of price-moving information

Model Constraints:

- Linear classifier cannot capture complex nonlinear patterns
- F1-scores insufficient for profitable trading after transaction costs
- Some assets (Polkadot, Avalanche) show weak signals requiring custom handling

Name: Manan Amit Shah

Methodological Constraints:

- In-sample results only; true out-of-sample testing on future data not available
- Assumes technical indicator patterns remain stationary
- Limited grid search scope due to computational constraints

Critical Challenges Addressed

1. **Data source instability:** Switched from CoinGecko (unreliable) to Yahoo Finance; standardized column naming
2. **Insufficient data:** Extended window to 60 days for indicator warm-up; modeled only on recent 30 days
3. **Zero metrics:** Reduced features from 21 to 8 for volatile assets; increased training windows to 16-18 days
4. **Model selection:** Tested LR, decision trees, ensembles; logistic regression proved most stable
5. **Validation approach:** Switched from fixed rolling windows to expanding windows for stability
6. **Class imbalance:** Applied balanced class weights; adjusted significant-move thresholds per asset

Assumptions and Risks

1. **Technical indicators are stationary:** Breaks during market regime changes
2. **Past patterns predict future:** Cryptocurrency markets frequently violate this assumption
3. **Equal misclassification costs:** Real trading has asymmetric cost structures
4. **One-day prediction is achievable:** Financial literature suggests inherent difficulty

Conclusions

The model demonstrates that next-day cryptocurrency direction exhibits weak but consistent predictability using technical indicators alone. F1-scores of 0.30-0.45 are above random but insufficient for reliable trading. The work validates academic findings that short term financial prediction is inherently difficult. For practical application, improvements would require: (1) integration with on-chain and sentiment data, (2) longer prediction horizons (weekly vs. daily), (3) regime-aware models, and (4) comprehensive backtesting with realistic transaction costs.

Recommendations for Extension

- Incorporate on-chain metrics, social sentiment, and macroeconomic indicators
- Experiment with tree-based ensembles (XGBoost) or neural networks (LSTM)
- Extend prediction horizon to weekly or monthly scales
- Implement full trading simulation with position sizing and risk management
- Deploy production pipeline with automated retraining and drift monitoring

Name: Manan Amit Shah

References

- Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25(2), 383-417.
- Malkiel, B. G. (1973). A Random Walk Down Wall Street. W.W. Norton & Company.
- Lo, A. W., & MacKinlay, A. C. (1988). Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test. The Review of Financial Studies, 1(1), 41-66.
- Wilder, J. W. (1978). New Concepts in Technical Trading Systems. Trend Research Limited.
- Murphy, J. J. (1999). Technical Analysis of the Financial Markets. Prentice Hall.
- Appel, G. (1979). The Moving Average Convergence-Divergence Trading Method. Signalert Corp.
- Bollinger, J. (1992). Bollinger on Bollinger Bands. McGraw-Hill.
- McNally, S., Roche, J., & Caton, S. (2018). Predicting the Price of Bitcoin Using Machine Learning. In 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP) (pp. 339-343). IEEE.
- Kristjanpoller, W., & Minutolo, M. C. (2016). Price Jump Prediction in a Regulated Power Market by Means of Recurrent Neural Networks. Applied Energy, 175, 313-321.
- Greaves, A., & Au, B. (2015). Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin. arXiv preprint arXiv:1509.01149.
- Cerqueira, V., Torgo, L., & Smiljanić, I. (2019). Evaluating Time Series Forecasting Models: An Empirical Study on Performance Estimation Methods. arXiv preprint arXiv:1905.11286.
- Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nd ed.). OTexts.
- Bergmeir, C., & Benítez, J. M. (2012). On the Use of Cross-Validation for Time Series. Information Sciences, 191, 192-213.
- Leitch, G., & Tanner, J. E. (1991). Evaluation of Forecasts. Handbook of Statistics, 9, 207-234.
- Campbell, J. Y. (1987). Stock Returns and the Term Structure. Journal of Financial Economics, 18(2), 373-399.
- Pesaran, M. H., & Timmermann, A. (1995). Predictability of Stock Returns: Robustness and Economic Significance. The Journal of Finance, 50(4), 1201-1228.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.
- Ng, A. (2019). Machine Learning Yearning. Technical Strategy for AI Engineers and Managers.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer.
- Bariviera, A. F. (2018). The Inefficiency of Bitcoin Revisited: A Dynamic Approach. Economics Letters, 161, 1-4.
- Urquhart, A. (2016). The Inefficiency of Bitcoin. Economics Letters, 148, 80-82.
- Hou, K., Xue, C., & Zhang, L. (2015). Digesting Anomalies: An Investment Approach. Journal of Financial Economics, 98(2), 175-194.