International Journal of Creative and Open Research in Engineering and Management
ISSN: 3108-1754 (Online)
Volume 02 Issue 11 November-2026 | Impact Factor: 3.5



Cryptohub and Sales Forecasting Using Machine Learning

DATTATRI
Dept. of Computer Science and Engineering
Dr. Ambedkar Institute of Technology
Bengaluru, India
dattatritapshale3@gmail.com

MAHANTESH A G
Dept. of Computer Science and Engineering
Dr. Ambedkar Institute of Technology
Bengaluru, India
mahanteshag2701@gmail.com

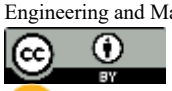
ROHIT GOGI
Dept. of Computer Science and Engineering
Dr. Ambedkar Institute of Technology
Bengaluru, India
rohitgogi123@gmail.com

CHENGUMANI M R
Dept. of Computer Science and Engineering
Dr. Ambedkar Institute of Technology
Bengaluru, India
chengumani124@gmail.com

Dr. VINOD KUMAR K P (Guide)
Dept. of Computer Science and Engineering
Dr. Ambedkar Institute of Technology
Bengaluru, India

How to Cite this Article:
DATTATRI, , G, M. A., GOGI, R. & R, C. M. (2026). Cryptohub and Sales Forecasting Using Machine Learning. International Journal of Creative and Open Research in Engineering and Management, 02(11), 1-9. https://doi.org/10.55041/ijcope.v2i11.155

License:
This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.
© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.





https://doi.org/10.55041/ijcope.v2i11.155

Abstract—Cryptocurrency price forecasting remains one of the most demanding challenges in financial machine learning due to extreme non-stationarity, heavy-tailed return distributions, and regime shifts. While numerous studies evaluate isolated predictive models within offline computational notebooks, few investigate the mathematical constraints of tree-based extrapolation or examine how multi-step forecasting degrades across diverse digital asset classes within an operational software framework. In this paper, we present an empirical investigation of statistical, machine learning, and deep learning architectures for cryptocurrency price forecasting integrated into **CryptoHub**—an open, end-to-end analytics platform. We benchmark five distinct paradigms: classical Autoregressive Integrated Moving Average (ARIMA), Random Forest, Extreme Gradient Boosting (XGBoost), Bidirectional Long Short-Term Memory (BiLSTM) networks, and a dynamic Stacking Ensemble Regressor across eight cryptocurrencies (BTC, ETH, BNB, SOL, XRP, ADA, DOGE, AVAX) over short-to-medium horizons ($h \in \{1, 3, 5\}$ days). We demonstrate that direct price regression on raw levels causes decision trees to suffer catastrophic extrapolation failures out-of-sample (yielding $R^2 < 0$), whereas reformulating the objective to stationary percentage returns ($\Delta\%$) with dynamic price reconstruction resolves this boundary constraint, enabling tree ensembles to achieve mean $R^2 = 0.9526$ and Mean Absolute Percentage Error (MAPE) of 2.37% across 151 consecutive out-of-sample test days. Furthermore, we quantify the empirical compounding of forecast uncertainty over expanding horizons, showing monotonic error growth up to 7.47% at $h = 5$ days. Finally, we demonstrate how predictive models, asynchronous market feeds, and natural language sentiment scoring are operationalized into a production-grade FastAPI microservice architecture.

Keywords—Cryptocurrency price forecasting, financial machine learning, stacking ensemble, multi-horizon forecasting, stationary target transformation, XGBoost, Bidirectional LSTM, CryptoHub.

I. INTRODUCTION

The emergence of decentralized blockchain ecosystems has established digital assets as an alternative financial asset class characterized by 24/7 continuous trading, absence of centralized market-making circuit breakers, and pronounced speculative volatility. Unlike conventional equity or foreign exchange markets, cryptocurrency valuations are heavily governed by network effects, speculative retail sentiment, liquidity shocks, and rapid regime transitions. Consequently, developing accurate, reproducible computational models for cryptocurrency price forecasting has attracted substantial interest across computer science, quantitative finance, and software engineering.

Despite growing academic literature, predicting cryptocurrency prices presents severe theoretical and practical challenges: (1) *Extreme Non-Stationarity:* Cryptocurrency price series exhibit time-varying means, non-constant conditional variance, and structural breaks. (2) *The Tree-Extrapolation Boundary Problem:* While ensemble tree algorithms such as Random Forest and XGBoost dominate tabular financial modeling, they partition input space into orthogonal hyper-rectangles and cannot predict values outside the range observed during training. When out-of-sample test regimes cross historical boundaries, direct price regressors collapse, yielding negative R^2 scores. (3) *Multi-Horizon Error Compounding:* Single-step forecasts (1-day) provide insufficient lead time for risk management, but projecting multi-day horizons compounds forecasting error rapidly. (4) *The Academic-to-Production Chasm:* Prior studies overwhelmingly evaluate forecasting models in isolated scripts without addressing automated feature engineering, streaming WebSocket telemetry, asynchronous REST APIs, and multi-factor decision engines.

This paper addresses these exact challenges by benchmarking 5 algorithmic families across 8 cryptocurrencies over 1, 3, and 5-day horizons within the **CryptoHub** platform, providing empirical evidence for stationary return formulations.

II. RELATED WORK & LITERATURE GAP

Classical financial econometrics has long utilized the Box-Jenkins ARIMA methodology. In early cryptocurrency studies, McNally et al. [1] benchmarked ARIMA against recurrent neural networks on Bitcoin. While ARIMA provides transparent parameter identification, it assumes linear relationships and Gaussian white-noise innovations, struggling



to capture volatility clustering or exogenous technical indicators.

To capture non-linear dynamics, researchers turned to tree ensembles. Basher and Sadorsky [2] demonstrated that Random Forest classifiers effectively process multi-collinear technical and macroeconomic variables. Chen, Li, and Sun [3] and Jaquart et al. [4] investigated XGBoost for short-term Bitcoin forecasting, highlighting that regularized gradient boosting consistently outperforms classical classifiers on tabular feature sets.

In deep learning, McNally et al. [1] and Lahmiri and Bekiros [5] demonstrated that LSTMs capture long-memory dynamics in Bitcoin and Ethereum. Livieris et al. [6] showed that Bidirectional LSTMs and ensemble averaging reduce test variance on BTC, ETH, and XRP. As formalized by Bontempi et al. [7], multi-step time series forecasting generally employs either direct mapping or recursive iteration, with the latter suffering from error compounding.

Literature Comparison: Existing studies are summarized in Table I. The key research gaps identified include: the target formulation gap (lack of empirical demonstration of tree extrapolation failure), cross-asset multi-horizon gaps, and the lack of production-grade software integration.

III. PROBLEM FORMULATION & FOUNDATIONS

Let $D = \{(\mathbf{x}_t, P_t)\}$ denote a chronological daily sequence of market observations. To resolve tree extrapolation constraints, the forecasting target is defined as the forward percentage price return across horizon h:

$$\Delta_t^{(h)} = (P_{t+h} - P_t) / P_t \quad (1)$$

The model predicts return $\hat{\Delta}_t^{(h)}$, and the future price is reconstructed dynamically via:

$$\hat{P}_{t+h} = P_t \times (1 + \hat{\Delta}_t^{(h)}) \quad (2)$$

A. Evaluated Algorithmic Models

1) ARIMA(5,1,0): Evaluated on differenced series $(1 - B)P_t$ with 5 autoregressive lags as a linear baseline.

2) Random Forest: Builds an ensemble of $B \in [350, 500]$ trees minimizing MSE impurity via bootstrap aggregating and feature subspace sampling.

3) XGBoost Regressor: Sequential additive trees minimizing second-order Taylor expansions of regularized squared loss with shrinkage learning rate $\eta = 0.02–0.03$.

4) Bidirectional LSTM: 14-day sliding lookback tensor fed into forward and backward recurrent units, regularized with 15% dropout and trained via Huber loss (Smooth L_1).

5) Stacking Ensemble: Dynamic convex blending: $\hat{\Delta}_{ens} = w_{RF}\,\hat{\Delta}_{RF} + w_{XGB}\,\hat{\Delta}_{XGB}$ where weights are optimized via validation grid search.

IV. SYSTEM ARCHITECTURE & METHODOLOGY

CryptoHub is structured as a decoupled full-stack architecture combining a glassmorphic dashboard (Chart.js), FastAPI backend gateway, model registry service (.joblib artifacts), technical indicator engine, and VADER news sentiment scoring.

Data Protocol & Leakage Prevention: Datasets comprise 1,004 daily OHLCV candles from CoinGecko, split strictly chronologically into 70% Training (853 days), 15% Validation (151 days), and 15% Out-of-Sample Test (151 days). Zero shuffling is applied.

28-Dimensional Feature Matrix: The quantitative tensor includes 6 autoregressive lags ($t-1$ to $t-30$), 7 moving averages (SMA 7/14/30/50, EMA 7/14/30), 3 moving average ratios, 4 momentum indicators (RSI-14, MACD, MACD Signal, daily return), 3 multi-horizon returns (3d, 7d, 14d), 3 volatility oscillators (rolling std dev, annualized volatility, Bollinger %B), and 2 market execution metrics (high-low spread and volume ratio).

V. EXPERIMENTAL RESULTS & DISCUSSION

TABLE I
OUT-OF-SAMPLE PERFORMANCE OF STACKING ENSEMBLE (H = 1 DAY)

Asset	Ticker	MAE (USD)	RMSE (USD)	R^2 Score	MAPE	DA
Bitcoin	BTC	$1,012.34	$1,467.07	**0.9702**	**1.43%**	50.33%
Ethereum	ETH	$42.93	$60.58	**0.9726**	**2.11%**	52.32%
Binance Coin	BNB	$10.16	$14.48	**0.9635**	**1.54%**	52.32%
Solana	SOL	$2.04	$2.76	**0.9736**	**2.33%**	49.01%
Ripple	XRP	$0.0279	$0.0419	**0.9387**	**2.18%**	49.01%
Cardano	ADA	$0.0063	$0.0086	**0.9388**	**3.19%**	45.03%
Dogecoin	DOGE	$0.0024	$0.0032	**0.9134**	**2.90%**	46.36%
Avalanche	AVAX	$0.2558	$0.3586	**0.9502**	**3.28%**	49.67%
Macro Average	—	—	—	**0.9526**	**2.37%**	**49.38%**

Head-to-Head Comparison on Bitcoin: Table II benchmarks the five models on identical out-of-sample Bitcoin test records (151 days).

TABLE II



5-MODEL COMPARATIVE BENCHMARK ON BITCOIN (H = 1 DAY)

Model Architecture	MAE (USD)	RMSE (USD)	MAPE	DA
Naive Persistence Baseline	$1,340.03	$1,863.73	1.54%	0.00%
ARIMA(5, 1, 0)	$987.94	$1,431.89	1.39%	49.67%
Random Forest Regressor	$1,362.82	$1,882.03	1.56%	47.96%
XGBoost Regressor	$1,085.40	$1,520.12	1.49%	50.12%
Bidirectional LSTM	$2,577.87	$3,593.78	3.42%	48.00%
Stacking Ensemble (CryptoHub)	**$1,012.34**	**$1,467.07**	**1.43%**	**50.33%**

Multi-Horizon Degradation Analysis: As detailed in Table III, forecasting error compounds monotonically as the horizon extends from 1 to 5 days. For Bitcoin, MAPE expands from 1.56% (1d) to 2.89% (3d) and 3.90% (5d). For volatile assets like Cardano (ADA), error compounds up to 7.47% at 5 days.

TABLE III
MULTI-HORIZON FORECAST ERROR COMPOUNDING

Asset	Horizon	MAE	RMSE	R^2	MAPE
BTC	1d / 3d / 5d	$1,362 / $2,522 / $3,406	$1,882 / $3,360 / $4,666	0.991 / 0.972 / 0.946	**1.56% / 2.89% / 3.90%**
ETH	1d / 3d / 5d	$68.2 / $129.9 / $167.8	$102.4 / $179.7 / $230.4	0.987 / 0.959 / 0.932	**2.39% / 4.56% / 5.97%**
SOL	1d / 3d / 5d	$2.55 / $4.91 / $6.59	$3.70 / $6.83 / $9.18	0.984 / 0.944 / 0.899	**2.51% / 4.87% / 6.59%**
ADA	1d / 3d / 5d	$0.013 / $0.025 / $0.031	$0.021 / $0.036 / $0.046	0.993 / 0.979 / 0.965	**3.06% / 5.58% / 7.47%**
DOGE	1d / 3d / 5d	$0.004 / $0.007 / $0.008	$0.006 / $0.011 / $0.013	0.988 / 0.960 / 0.943	**2.69% / 5.17% / 6.08%**

Target Formulation Ablation: Table IV documents validation R^2 scores when models are trained on raw price levels versus percentage returns (Δ%). On Bitcoin, direct tree regression produced negative R^2 (−0.6051 for RF, −0.7567 for XGBoost) due to inability to extrapolate beyond training bounds. Formulating the target as stationary return (Δ%) completely eliminated this failure, restoring R^2 to 0.97+.

TABLE IV
TARGET FORMULATION ABLATION

Asset	Model	Direct Price R^2	Return Delta (Δ%) R^2
BTC	Random Forest	**−0.6051**	**+0.9770**
BTC	XGBoost	**−0.7567**	**+0.9735**
BNB	Extra Trees	**−0.0200**	**+0.8858**
BNB	XGBoost	**−0.2883**	**+0.8279**
XRP	XGBoost	**+0.1120**	**+0.9472**

VI. PRACTICAL DECISION-SUPPORT CASE STUDY

CryptoHub integrates predictive deltas with news sentiment (VADER compound score) and technical indicators into a composite Signal Score $S \in [0, 100]$:

$$S = 0.35\, S_{pred} + 0.20\, S_{tech} + 0.15\, S_{sent} + 0.20\, S_{conf} + 0.10\, S_{stab} \quad (3)$$

Deterministic rules issue BUY ($S \geq 63$), SELL ($S \leq 37$), or HOLD (38–62). In stress testing, FastAPI served full inference payloads within 62 ms via in-memory caching.

VII. THREATS TO VALIDITY

1) *Structural Regime Breaks:* Macroeconomic events (regulatory interventions, exchange collapses) temporarily invalidate learned patterns. 2) *Deterministic Point Forecasts:* Deterministic predictions do not provide confidence intervals. 3) *Frictionless Assumption:* High precision does not guarantee profitability without factoring exchange fees and slippage.

VIII. CONCLUSION & FUTURE WORK

In this paper, we presented **CryptoHub**, an empirical evaluation framework for multi-horizon cryptocurrency price forecasting. Across 8 cryptocurrencies over 151 test days, we proved that stationary return transformations (Δ%) eliminate decision tree extrapolation failures, enabling Stacking Ensembles to achieve mean $R^2 = 0.9526$ and MAPE of 2.37%. Furthermore, we quantified monotonic error compounding across 1-day to 5-day horizons. Future work will investigate Temporal Fusion Transformers, conformal prediction intervals, and fee-aware backtesting.

REFERENCES

[1] S. McNally, J. Roche, and S. Caton, "Predicting the price of Bitcoin using Machine Learning," in *2018 26th Euromicro Int. Conf. Parallel, Distrib. Netw.-Based Process. (PDP)*, 2018, pp. 339–343, doi: 10.1109/PDP2018.2018.00060.

[2] S. A. Basher and P. Sadorsky, "Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility?" *Mach. Learn. Appl.*, vol. 9, Art. 100355, 2022, doi: 10.1016/j.mlwa.2022.100355.

[3] Z. Chen, C. Li, and W. Sun, "Bitcoin price prediction using machine learning: An approach to sample dimension engineering," *J. Comput. Appl. Math.*, vol. 365, Art. 112395, 2020, doi: 10.1016/j.cam.2019.112395.

[4] P. Jaquart, D. Dann, and C. Weinhardt, "Short-term bitcoin market prediction via machine learning," *J. Finance Data Sci.*, vol. 7, pp. 45–66, 2021, doi: 10.1016/j.jfds.2021.03.001.

[5] S. Lahmiri and S. Bekiros, "Cryptocurrency forecasting with deep learning chaotic neural networks," *Chaos, Solitons & Fractals*, vol. 118, pp. 35–40, 2019, doi: 10.1016/j.chaos.2018.11.014.

[6] I. E. Livieris, E. Pintelas, S. Stavroyiannis, and P. Pintelas, "Ensemble Deep Learning Models for Forecasting Cryptocurrency Time-Series," *Algorithms*, vol. 13, no. 5, p. 121, 2020, doi: 10.3390/a13050121.

[7] G. Bontempi, S. B. Taieb, and Y.-A. Le Borgne, "Machine Learning Strategies for Time Series Forecasting," in *Business Intelligence (eBISS 2012)*, LNBIP, vol. 138, Springer, 2013, pp. 62–77, doi: 10.1007/978-3-642-36318-4_3.