*Corresponding Author's Email: nahla.smida@gmail.com
Proceedings of the International Conference on New Trends in Management, Business and Economics
Vol. 3, Issue. 1, 2025, pp. 34-47
DOI: https://doi.org/10.33422/icnmbe.v3i1.1356

ISSN: 3030-0347 online

Forecasting Cryptocurrency Market Trends Using Machine Learning on Multidimensional Time-Series Data

Michal Koren[1*], Steven Berger[2], and Raz Levy[2]

[1]School of Industrial Engineering and Management, Shenkar-Engineering. Design. Art, Ramat Gan 5252626, Israel

[2]Faculty of Engineering, Bar-Ilan University, Ramat Gan 5290002, Israel

Abstract

The cryptocurrency market is characterized by extreme volatility and strong behavioral biases, which often lead to irrational decision-making by both retail and institutional traders. To address this challenge, we present a systematic, data-driven forecasting framework that combines classical machine learning and deep learning models to predict short-term price direction in the Bitcoin market. Our approach leverages a decade of hourly Bitcoin price data and constructs a rich set of engineered features across four key domains: momentum (rate of change), candlestick psychology, volatility patterns, and volume dynamics. These features capture both structural and behavioral signals that are critical in high-frequency trading environments.

We evaluate ten supervised learning models including tree-based ensembles, logistic regression, recurrent and convolutional networks, and hybrid architectures—on their ability to forecast 4-hour price movements. To enhance robustness, we implement multiple ensemble techniques: majority voting, weighted soft voting, and rule-based threshold combinations. Experimental results show that ensemble models significantly outperform individual classifiers. Weighted soft voting achieves a precision of 0.6375, while rule-based ensembles reach over 0.85 precision with ultra-low activity, making both methods suitable for high-confidence entry signals. By aligning predictive accuracy with trading practicality, our framework provides a modular and emotion-agnostic decision-support system for cryptocurrency trading. The results demonstrate the potential of combining structured feature engineering with model diversity to navigate the complexities of a highly volatile, behavior-driven market.

Keywords: Trend Forecasting; Machine Learning; Crypto; Bitcoin

1. Introduction

The emergence of cryptocurrencies has significantly reshaped the global financial landscape, introducing a decentralized, always-on, and highly volatile asset class (Baur and Oll, 2024). Leading this transformation is Bitcoin, which serves concurrently as a speculative investment, a hedge against traditional financial systems, and a digital store of value. Unlike conventional equities, cryptocurrency markets are marked by extreme price volatility (Bouri et al., 2024), limited regulatory oversight, and pronounced information asymmetry—factors that pose distinct challenges for traders and analysts alike.

In this environment, emotion-driven decision-making frequently overshadows rational, strategic thinking (Song, 2025). Estimates suggest that up to 90% of retail cryptocurrency traders experience financial losses, largely due to cognitive biases such as fear of missing out (FOMO), overconfidence, and herd behavior (Vidani, 2024). These behavioral tendencies reduce the effectiveness of discretionary trading and highlight the need for systematic, data-driven approaches. While traditional econometric models such as ARIMA, GARCH, and other linear time-series techniques have proven useful in structured financial contexts, they often fall short in capturing the nonlinear dynamics, non-stationarity, and sudden regime shifts that characterize cryptocurrency markets (Berger and Koubová, 2024). In response, researchers have increasingly turned to machine learning (ML) and deep learning (DL) methods, which are well equipped to uncover patterns in high-frequency, high-dimensional data without relying on rigid parametric assumptions.

This paper presents a robust, modular forecasting framework aimed at predicting short-term directional movements in Bitcoin's price. The main contributions of this paper are as follows: (1) a comprehensive comparative analysis of supervised learning models, for forecasting directional price movements; (2) an investigation into ensemble strategies, such as majority voting and precision-weighted soft voting, to evaluate whether combining heterogeneous model predictions enhances predictive robustness, precision, and generalizability; and (3) the introduction of the trades per month (TPM) metric, which adds a practical dimension to model evaluation. While the academic literature often emphasizes precision, recall, or F_1 score in isolation, TPM reflects how frequently a model engages the market, a critical factor for portfolio construction, capital deployment, and strategy calibration. This alignment between statistical performance and real-world applicability enhances both the interpretability and utility of the proposed framework.

To illustrate the complex and nonlinear characteristics of the data, Figure 1 presents Bitcoin's historical price trajectory from 2015 to 2025 based on data from the Gemini cryptocurrency exchange. The plot highlights recurrent boom-and-bust cycles, reinforcing the need for flexible, nonparametric models that can accommodate volatility clustering and structural regime shifts.

Figure 1: Bitcoin's historical price trajectory from 2015 to 2025
Source: data from https://www.gemini.com/prices/bitcoin

By addressing both the cognitive biases inherent in discretionary trading and the limitations of traditional econometric models, this study bridges the domains of behavioral finance and predictive analytics. The proposed framework is scalable across various timeframes and asset classes, laying a foundation for future research in algorithmic trading, risk management, and the development of quantitative strategies for crypto assets.

Understanding the psychological underpinnings of investor behavior has become an increasingly important focus in financial research, particularly through the lens of behavioral finance (Padmavathy, 2024). This is especially relevant in the context of cryptocurrency markets, which are characterized by high volatility, limited regulation, and 24/7 trading conditions that amplify emotional decision-making and speculative behavior. Agarwal and Tewari (2024) provide a comprehensive synthesis of psychological factors such as fear, greed, happiness, and overconfidence that frequently override rational judgment in stock market investments. These biases are reflected in well-documented phenomena such as the Disposition Effect, whereby investors tend to prematurely sell winning assets while holding onto losing ones. To mitigate such biases, the authors advocate for strategies including portfolio diversification, goal-setting, and disciplined risk management. However, they emphasize that the majority of empirical studies to date have concentrated on traditional financial markets, leaving the behavioral dynamics of cryptocurrency markets relatively underexplored.

Addressing this gap, Schatzmann and Haslhofer (2023) examined whether similar behavioral biases exist among Bitcoin investors. Using granular blockchain transaction data spanning 2013 to 2021, they found strong evidence of irrational trading behaviors, particularly during periods of heightened market sentiment, such as the 2017 bull market. Despite Bitcoin's decentralized and algorithmic infrastructure, investor behavior exhibited familiar emotional patterns, including loss aversion, herd behavior, and premature profit-taking. Notably, the prevalence of these biases appeared to decline in more recent years, suggesting a possible learning curve among retail investors and signs of increasing market maturity.

Collectively, these findings underscore a critical insight: in emotionally charged and highly volatile environments such as cryptocurrency markets, human decision-making is especially prone to error. This has spurred growing interest in algorithmic and data-driven approaches designed to reduce the psychological noise inherent in discretionary trading. In this context, ML and DL techniques present compelling alternatives. Unlike traditional econometric models, ML and DL methods are well suited to capturing nonlinear dependencies, temporal irregularities, and high-dimensional feature interactions—hallmarks of modern crypto market behavior.

For instance, Tran Phuoc et al. (2024) employed long short-term memory (LSTM) networks to forecast stock price trends in the Vietnamese market using technical indicators such as the simple moving average (SMA), relative strength index (RSI), and moving average convergence/divergence (MACD). Their model achieved an accuracy exceeding 93%, though the authors cautioned that performance is highly context dependent and may not generalize across different asset classes or market structures. Similarly, Basher and Sadorsky (2022) demonstrated the effectiveness of random-forest models in forecasting Bitcoin and gold price directions using macroeconomic indicators, including interest rates and oil volatility (OVX). Their ensemble-based approach outperformed traditional logistic regression models, highlighting the advantages of nonparametric methods in capturing complex market dynamics.

Shahariar and Akhand (2023) applied extreme gradient boosting (XGBoost) for stock price classification based on a set of technical indicators. Their findings affirmed the utility of decision-tree ensembles in modeling nonlinear feature interactions, while also emphasizing the potential benefits of incorporating sentiment analysis and real-time data streams. Most notably, Bistarelli et al. (2024) conducted a comprehensive meta-analysis of 100 cryptocurrency forecasting studies, reporting average precision rates of 55%-62%, with top-performing models achieving up to 70%. These results reinforce both the promise and the limitations of current ML methodologies, and they underscore the need for continued innovation, particularly in the development of hybrid and ensemble strategies, which constitute a central focus of the present paper.

2. Material and Methods

This section outlines the entire methodological framework used to develop, train, and evaluate predictive models for short-term forecasting of Bitcoin price direction. As this is a binary classification problem, the target variable indicates whether the price will increase (1) or not (0) in the following 4-hour interval. We selected the 4-hour interval to strike a balance between signal stability and trading frequency. Whereas most prior research focuses on daily horizons, the 4-hour timeframe captures meaningful intraday price dynamics, while avoiding excessive exposure to extreme, short-lived market distortions, such as those triggered by tweets or sudden news events. Compared to daily data, it offers a greater number of actionable trading signals, enabling more opportunities for executing strategy without being overwhelmed by the noise inherent in ultra-short intervals (e.g., 1-hour or shorter). We integrate systematic feature engineering, a diverse array of machine learning and deep learning architectures, and robust ensemble techniques to enhance generalizability and reduce overfitting.

2.1. Data Collection and Distribution Analysis

The foundational dataset consisted of hourly Bitcoin Open, High, Low, Close, Volume (OHLCV) data spanning a decade (2015-2025). These data were aggregated into authentic 4-hour candles to better capture intraday market dynamics and trading session effects. For each 4-hour window, we derived a comprehensive set of features grounded in financial theory and trader psychology, grouped into four conceptual categories:

- **Rate of Change (ROC):** Momentum-based indicators capturing short- and long-term directional strength.
- **Candlestick Psychology:** Features based on classic price-action heuristics (e.g., body size, shadow ratios).
- **Volatility Awareness:** Rolling standard deviations, Bollinger bands, and normalized ranges.
- **Volume Dynamics:** Indicators reflecting market participation, liquidity flows, and volume-price divergence.

These features were constructed using rolling statistics, ratio encodings, and multi-scale differentials. The final dataset comprised over 80 features. To prevent multicollinearity and overfitting, model-specific feature selection was applied prior to training.

Before proceeding with modeling, we conducted an exploratory distributional analysis of Bitcoin's closing prices. Many classical time-series models (e.g., ARIMA, VAR) assume normality or near-normality of inputs. To assess whether this assumption held, we applied common transformations—including square root, logarithmic, and Box-Cox ($\lambda = 0.23$)—to the price series. As illustrated in Figure 2, none of these transformations yielded a unimodal, symmetric distribution. The data remained heavy-tailed and multimodal, reinforcing the need to move beyond parametric techniques. Consequently, we focused our efforts on nonparametric models, such as decision tree ensembles and deep neural networks, that are inherently robust to skewness, kurtosis, and other non-Gaussian traits.

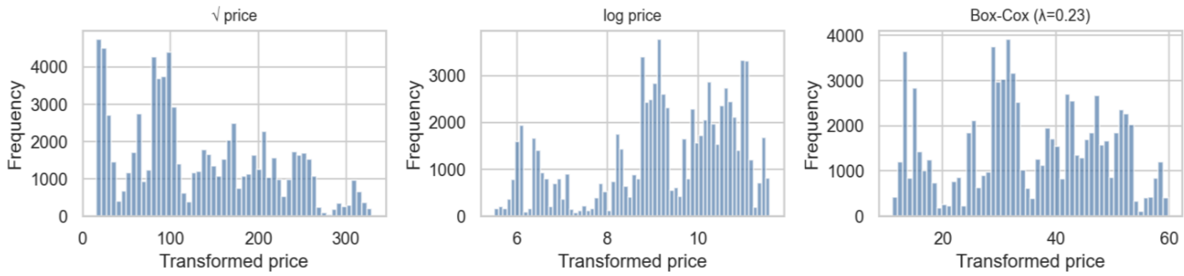

Figure 2: Distribution of transformed Bitcoin closing prices: square root (left), logarithmic (middle), and Box-Cox transformations (right).

2.2. Model Selection

To capture a wide range of market dynamics and modeling biases, we employed ten supervised learning models spanning classical statistical methods, ensemble-based approaches, and deep neural network architectures. These models were selected for their capacity to model nonlinear relationships, temporal dependencies, and complex feature interactions, all prevalent in high-frequency cryptocurrency data. All models were trained on historical data and evaluated using a time-consistent cross-validation scheme to prevent look-ahead bias and data leakage.

1. **Random Forest (RF):** A bagging-based ensemble of decision trees, offering robust baseline performance, resistance to overfitting, and interpretability through feature importance scores.
2. **XGBoost (XGB):** A gradient-boosted decision tree framework, optimized using the Optuna hyperparameter tuning library; known for its predictive accuracy and robustness in noisy environments.
3. **LightGBM (LGBM):** A graphics processing unit (GPU)-accelerated gradient boosting method employing leaf-wise tree growth and dropout regularization, particularly well-suited for high-dimensional feature spaces.
4. **CatBoost (CB):** A boosting algorithm with native support for categorical features and built-in handling of class imbalance, yielding strong generalization in diverse settings.
5. **Logistic Regression (LR):** A linear classification baseline tested under multiple regularization schemes (L1, L2), with synthetic oversampling (SMOTE, ADASYN) and class-weight adjustments to address label imbalance.
6. **Long Short-Term Memory (LSTM):** A recurrent neural network architecture designed to capture sequential dependencies and long-term memory in time-series data.
7. **Gated Recurrent Unit (GRU):** A simplified alternative to LSTM, optimized for computational efficiency and precision, with tuning focused on the $F_{0.5}$ score to emphasize predictive precision.
8. **Convolutional Neural Network (CNN):** A deep convolutional model trained to detect local temporal patterns, volatility spikes, and short-term structural shifts in market behavior.
9. **CNN-LSTM Hybrid:** A composite architecture combining convolutional layers for spatial feature extraction with LSTM layers for sequential modeling.
10. **Temporal Convolutional Networks (TCN):** A class of deep models utilizing dilated causal convolutions and residual connections to capture long-range dependencies with parallelizable training.

All deep learning models were implemented on GPU-accelerated infrastructure and trained with dropout regularization, accelerated linear algebra (XLA) compilation for performance optimization, and hyperparameter tuning via Optuna to ensure fair and efficient model comparison. To enhance model robustness and capitalize on the complementary strengths of diverse architectures, we implemented and evaluated three ensemble strategies:

1. **Majority Voting:** This categorical ensemble method aggregates predictions by allowing each model to cast a vote for a specific class. The class receiving the most votes is selected as the final output, with all models contributing equally. This approach helps mitigate the influence of poorly performing models by relying on collective consensus (Sagi and Rokach, 2018).
2. **Weighted Soft Voting:** In this approach, each model's predicted probabilities are scaled based on its validation precision before aggregation. By assigning higher weights to more accurate models, this method emphasizes high-confidence predictions and improves stability near the classification threshold (Awan-Ur-Rahman, 2023).
3. **Rule-Based Ensemble:** This technique triggers predictions only when confidence scores from one or more models exceed predefined thresholds. By enforcing high-certainty conditions, this method prioritizes precision over recall and is particularly useful in risk-sensitive applications (McFadden, 2025).

We partitioned the dataset chronologically to maintain temporal integrity: training set (from January 2018 to September 2022), validation set (from October 2022 to October 2023) and test set (from November 2023 to March 2025). Traditional time-series cross-validation was avoided due to the risk of look-ahead bias. Instead, we used rolling holdouts and forward validation. All neural models were trained with early stopping, and final evaluation was conducted exclusively on unseen test data. Evaluation metrics included:

- **Precision:** Prioritized due to its alignment with profitable trade execution.
- **Recall:** Captures the model's sensitivity to potential upswings.
- **F_1 score:** Harmonic mean of precision and recall.
- **ROC Area Under the Curve (ROC AUC):** Measures the separability of the classes over varying thresholds.
- **Thresholds:** Thresholds were fixed at 0.5 unless otherwise optimized via validation tuning.

We employed the Optuna framework for automated hyperparameter optimization. The tree-structured Parzen estimator (TPE) sampler was used to efficiently explore high-dimensional spaces. Optimization objectives varied by model: standard F_1 was used for balanced evaluation and $F_{0.5}$ for precision-prioritized models like GRU and CNN variants.

Tuned parameters included learning rate, batch size, dropout rate, kernel size, L2 regularization strength, and network depth. Early pruning, memory management, and GPU utilization were employed to ensure scalable, efficient tuning. Best configurations were preserved for final training and downstream evaluation.

3. Results

To assess the baseline performance of each model in isolation, we evaluated predictive accuracy on the unseen test set (2023-2025), focusing on two primary metrics:

- **Precision:** The proportion of true positives among all upward predictions, reflecting the quality of trading signals.
- **Trades per Month (TPM):** A normalized operational metric derived from recall, estimating the average monthly frequency of triggered trades.

This dual-metric framework provides a practical lens for evaluating the tradeoff between signal quality (precision) and signal volume (TPM), which is critical in the context of real-world trading.

Figure 3 presents each model in precision-TPM space. Models positioned toward the upper left quadrant are both selective and accurate. The Random Forest model achieved the highest precision (0.585) with moderate trading frequency, offering a balanced, high-quality signal stream. CatBoost, LSTM, and LightGBM demonstrated strong tradeoffs between precision and activity, indicating suitability for medium-frequency strategies. GRU and TCN yielded the highest TPMs (>125), suggesting aggressive signal generation, albeit with lower precision.

Interestingly, Logistic Regression, despite its simplicity, delivered a precision of 0.551 with high trading frequency, showcasing the enduring utility of linear classifiers in well-engineered feature spaces.

These results illustrate the precision-recall tension inherent to financial classification problems: conservative models favor precision but trigger fewer trades, while exploratory models offer broader coverage at the cost of reliability. The choice of model should therefore be aligned with the desired strategic posture, whether risk-averse or opportunity-seeking (see full results in Appendix A, Table 1).

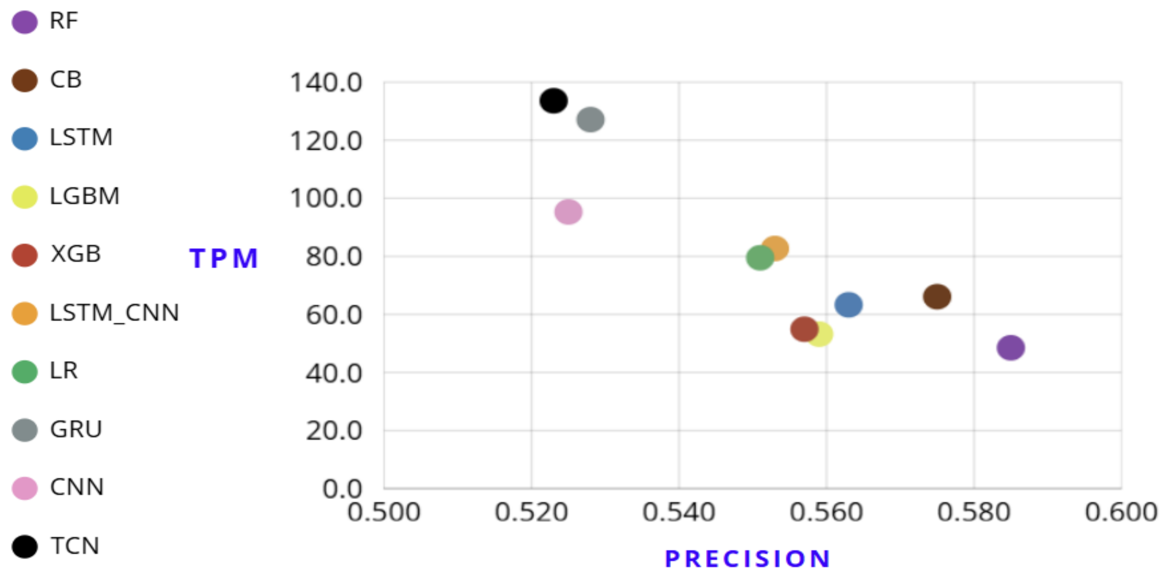

Figure 3: Precision and TPM scores for each model

3.1. Majority Voting

We evaluated hard voting ensembles, in which each constituent model casts a binary prediction (UP or DOWN) and the final output is determined by majority consensus. This approach is designed to reduce individual model variance and is particularly effective when aggregating predictions from diverse model architectures. The precision and TPM scores for the evaluated hard voting ensembles are summarized in Figure 4 (see full results in Appendix A, Table 2).

The top-performing ensemble, comprising GRU, LSTM, Random Forest, and CNN-LSTM, achieved a precision of 0.6157 and a TPM of 23.3, outperforming all individual models in both accuracy and market engagement. Several other combinations also demonstrated strong results. For instance, the ensemble of CNN, GRU, LSTM, and XGBoost achieved a precision of 0.6144 with a TPM of 27.5; CatBoost, GRU, LightGBM, and CNN-LSTM reached a precision of 0.6140 and TPM of 25.3; while GRU, LightGBM, LSTM, and Random Forest recorded a precision of 0.6113 and TPM of 28.5.

These findings underscore several important insights. First, architectural diversity appears to enhance generalization and improve ensemble robustness. Second, precision gains are consistent across a range of model configurations, indicating that performance is not overly sensitive to specific combinations. Finally, ensembles operating in the TPM range of 20-30 tend to strike a favorable balance between predictive confidence and market coverage, suggesting practical viability for active trading strategies. In summary, majority (hard) voting emerges as a computationally efficient and effective ensembling method that is particularly valuable in settings where robustness and model diversity are prioritized over fine-tuned calibration.

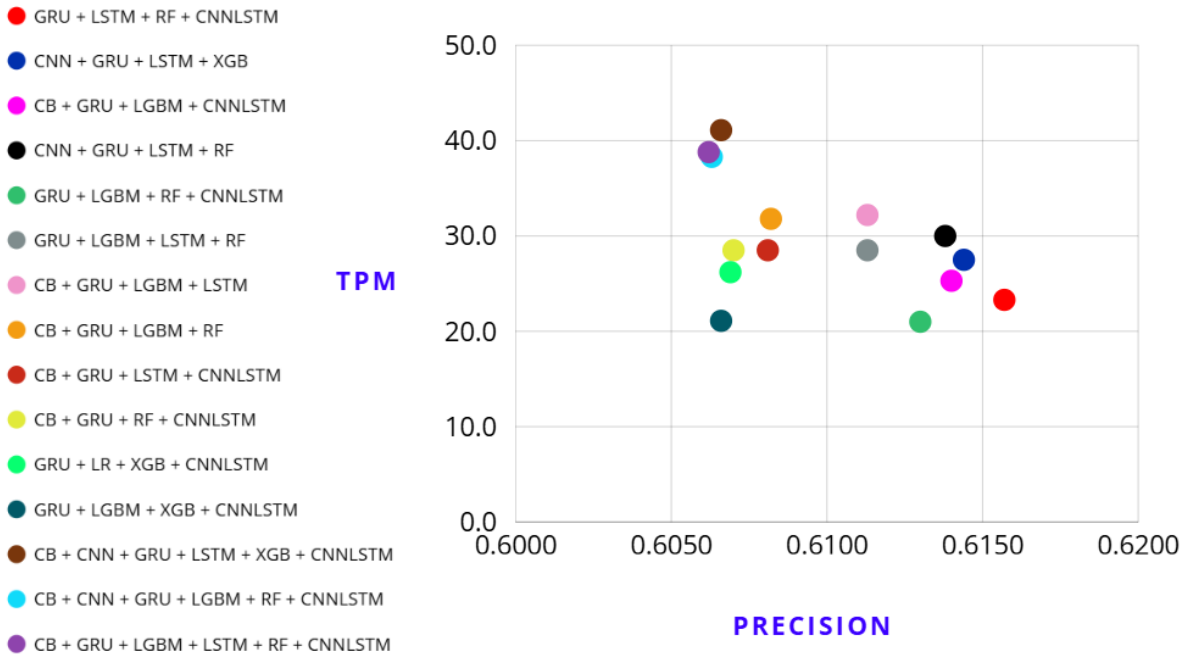

Figure 4: Precision and TPM scores for hard voting ensembles

3.2. Weighted Soft Voting

The second ensemble strategy we explored in this study is weighted soft voting, wherein predicted class probabilities from each model are combined through a linear weighting scheme. These weights are primarily derived from validation precision, allowing the ensemble to emphasize more reliable models while preserving probabilistic confidence. This method offers a flexible mechanism for balancing predictive performance with model trustworthiness.

The top-performing ensemble consisting of GRU and LightGBM achieved a precision of 0.6375 with a TPM of 4.2, representing a highly selective and low-activity configuration suitable for conservative trading strategies. Several other high-performing combinations also emerged. For example, the ensemble of GRU, Random Forest, and TCN reached a precision of 0.6312 with a TPM of 23.0; GRU, LightGBM, and TCN achieved 0.6304 precision and 4.8 TPM; GRU combined with Random Forest yielded 0.6288 precision and 22.3 TPM; while CatBoost, GRU, LightGBM, and TCN produced 0.6250 precision with a TPM of 13.2.

A common thread among the strongest ensembles is the inclusion of GRU and LightGBM, suggesting complementary modeling advantages: namely, GRU's capacity for learning temporal dependencies and LightGBM's strength in capturing structural feature interactions. Interestingly, expanding the ensemble beyond four models generally led to diminishing returns, reinforcing the principle that model quality and diversity outweigh sheer quantity in ensemble design.

As shown in Figure 5 (see full scores in Appendix A, Table 3), weighted soft voting demonstrates strong potential for fine-grained decision-making. Its tunability makes it especially attractive in portfolio management contexts, where dynamic capital constraints and shifting risk tolerances necessitate adaptable model behavior.

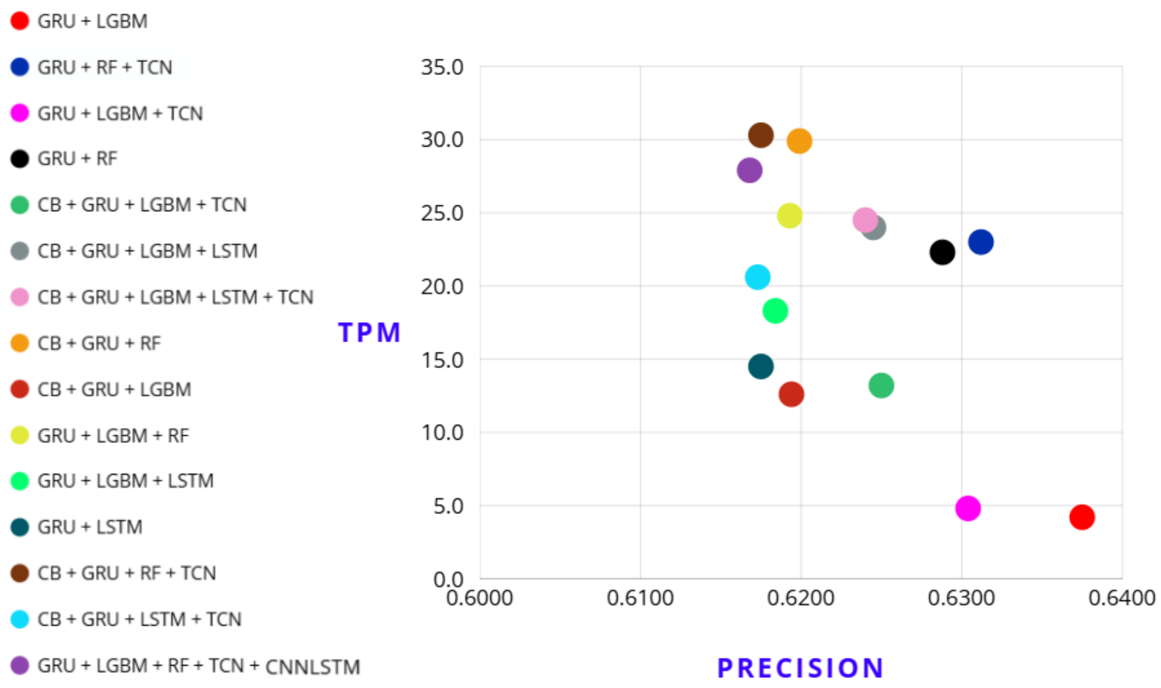

Figure 5: Precision and TPM scores for weighted soft voting ensembles

3.3. Rule-Based Threshold Combinations

The final ensembling strategy investigated in this study is rule-based threshold logic, which issues predictions only when specific compound conditions, typically based on model-level confidence, are jointly satisfied. Unlike soft voting, which aggregates probabilistic outputs, this approach enforces explicit agreement among high-confidence signals, thereby prioritizing precision over frequency.

As illustrated in Figure 6 (see full scores in Appendix A, Table 4), this method yielded the highest precision scores observed in the study. The most selective configuration, combining LightGBM and CNN-LSTM, achieved a precision of 0.8571 with a TPM of 0.4. Other top-performing combinations included CatBoost, LightGBM, XGBoost, and CNN-LSTM (precision 0.8333, TPM 0.3); CatBoost, LightGBM, LSTM, and CNN-LSTM (precision 0.8000, TPM 0.3); and a simpler ensemble of CatBoost and LightGBM (precision 0.7778, TPM 0.5). These ensembles effectively isolate high-confidence intersections between structurally diverse models, particularly boosted tree algorithms and sequence-based deep learning architectures.

While narrower rule sets produced the most accurate signals, expanding the ensemble to include five or more models improved trade frequency at the cost of some precision. For example, an ensemble consisting of CatBoost, GRU, LightGBM, XGBoost, and CNN-LSTM achieved a precision of 0.7059 with a TPM of 1.0. Similarly, CNN, LightGBM, LSTM, and Random Forest reached 0.6854 precision and 5.0 TPM, while the LSTM and XGBoost pair attained 0.6793 precision and 8.9 TPM.

Although these rule-based configurations are inherently low activity, their near-flawless entry signals make them particularly well suited for capital preservation strategies, as risk filters, or as components in signal confirmation layers within broader trading systems. Furthermore, their transparent and interpretable logic offers practical advantages in compliance-sensitive environments and decision-support systems where trader oversight or regulatory accountability is required.

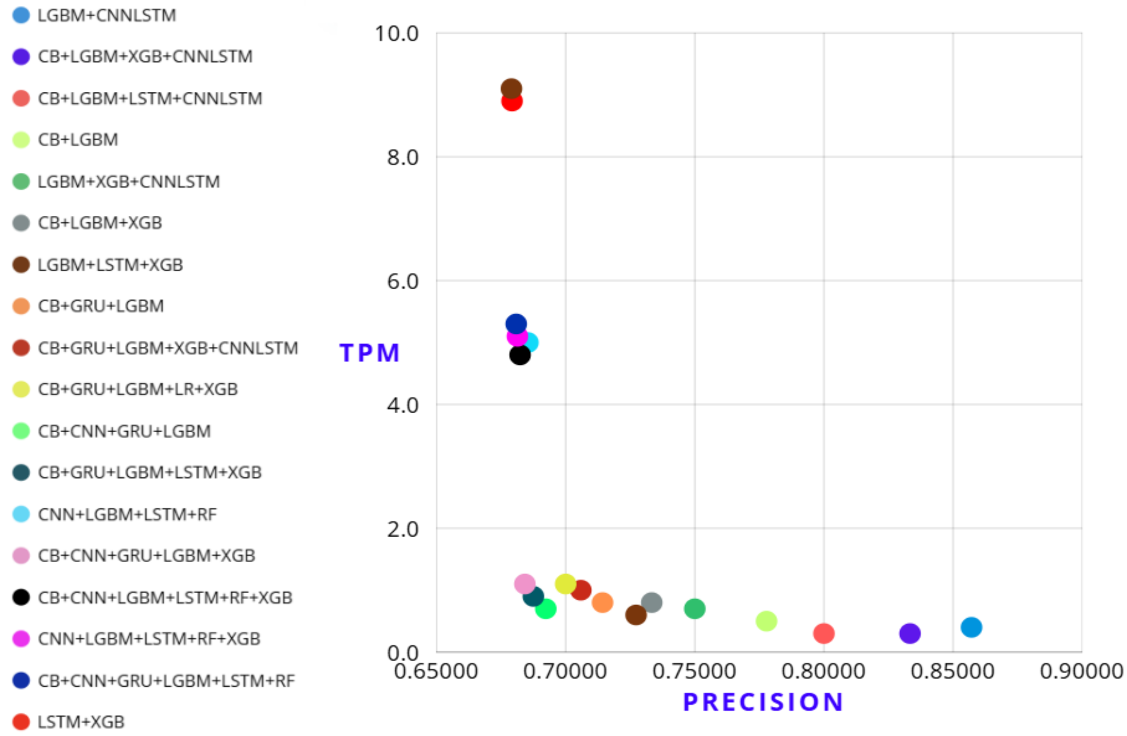

Figure 6: Precision and TPM scores for rule-based threshold combinations

3.4. Comparison between Methods

The ensemble strategies majority voting, weighted soft voting, and rule-based thresholding exhibit distinct performance profiles and operational tradeoffs, reflecting different strategic priorities in algorithmic trading. Majority voting ensembles, which aggregate binary decisions through a simple consensus mechanism, offered a strong balance between predictive precision and trade frequency. Top configurations, such as the GRU + LSTM + Random Forest + CNN-LSTM ensemble, achieved a precision of 0.6157 with a TPM of 23.3, outperforming all individual models. These ensembles benefit from architectural diversity and demonstrate consistent generalization, making them well suited for active strategies where robustness and coverage are key.

In contrast, weighted soft voting prioritized model reliability, using validation-based weights to combine probabilistic outputs. This method achieved the highest precision among scalable ensemble techniques, with the best-performing configuration (GRU + LightGBM) reaching a precision of 0.6375, though at a significantly lower TPM of 4.2. Such low-frequency, high-precision behavior makes soft voting particularly valuable for risk-sensitive trading systems, for which trade selectivity is paramount. Ensembles that included GRU and LightGBM consistently outperformed others, suggesting strong synergy between temporal and structural learning. The rule-based thresholding approach delivered the most precise predictions overall, with configurations such as LightGBM + CNN-LSTM achieving a precision of 0.8571 and TPM of just 0.4. However, this extreme selectivity comes at the cost of trade frequency. These ensembles are ideal for capital preservation strategies, signal confirmation layers, or compliance-aware environments where interpretability and precision outweigh volume. While broader rule sets improved TPM modestly, precision declined proportionally, reinforcing the precision-recall tradeoff inherent to this method.

4. Discussion

This paper introduces a comprehensive, modular framework for short-term directional forecasting of Bitcoin prices, which integrates classical machine learning and modern deep learning techniques and applies them to multidimensional, feature-engineered time-series data. By systematically incorporating indicators derived from market momentum, volatility regimes, candlestick psychology, and volume dynamics, the proposed approach is designed to capture the behavioral and structural intricacies of cryptocurrency markets.

A key insight emerging from the empirical results is the persistent tradeoff between predictive precision and trading frequency a well-documented tension in financial modeling. High-precision models, particularly those based on rule-based ensemble logic, consistently achieved near-expert signal quality (precision > 0.85) but exhibited very low activity (TPM < 1). In contrast, more aggressive architectures such as GRU, TCN, and CNN generated higher trading volumes but at the expense of reduced reliability, reflecting their exploratory and less conservative nature.

Among individual models, Random Forest and CatBoost demonstrated notable robustness and precision, outperforming other classical methods. Interestingly, Logistic Regression also yielded competitive results despite its simplicity, underscoring the influence of feature engineering over model complexity in certain scenarios. Within the deep learning domain, LSTM and CNN-LSTM hybrids performed reliably during periods of heightened market volatility, albeit with increased computational overhead. The most substantial performance improvements, however, were observed through the use of ensemble strategies, all of which outperformed their respective base learners:

- **Majority Voting:** Improved generalization by reducing variance across models, serving as a strong, low-complexity baseline.
- **Weighted Soft Voting:** Achieved the highest overall precision among scalable ensembles (up to 0.6375), enabling tunable trade-offs between signal conservatism and trading frequency.
- **Rule-Based Ensembles:** Produced highly selective signals with exceptional precision (>0.85), validating the utility of explicit logical constraints for generating high-confidence entry conditions.

Each ensemble mechanism aligns with distinct operational goals: rule-based ensembles are ideal for low-frequency, high-certainty strategies; weighted ensembles offer a balance between reliability and market engagement; and majority voting provides a flexible, plug-and-play solution for rapidly shifting environments.

Despite the promising results, several limitations should be acknowledged: (1) This study focuses exclusively on Bitcoin and a fixed 4-hour prediction window. Future research should evaluate the framework across multiple timeframes (e.g., 1-hour, daily) and additional digital assets to assess generalizability. (2) The current framework relies solely on historical price and volume data. Integrating alternative data sources including real-time order book dynamics, social media sentiment, and blockchain (on-chain) activity could provide a more comprehensive market view and improve predictive accuracy.

This study demonstrates the feasibility of constructing high-precision, emotion-agnostic trading models using structured price data alone. By integrating diverse model architectures with feature-rich representations into flexible ensemble systems, we present a blueprint for designing decision-support tools capable of navigating the volatility and non-stationarity inherent in cryptocurrency markets.

Among the ensemble approaches evaluated, weighted soft voting and rule-based thresholding emerged as particularly effective, each catering to different operational needs and risk-return profiles. These findings indicate that robust signal extraction is achievable even in highly speculative and behaviorally driven environments provided that model diversity, regularization, and decision calibration are carefully prioritized.

Ultimately, this work contributes both a benchmark and a modular foundation for future research in cryptocurrency forecasting. As digital asset markets continue to evolve in complexity and scale, the development of systematic, adaptive, and interpretable predictive frameworks will be essential in bridging the gap between raw market data and informed, algorithmically guided decision-making.

References

Agarwal, K., & Tewari, D. (2024). *Exploring emotional influences on investment choices: Stock selection and trading*. SSRN. https://doi.org/10.2139/ssrn.5114355

Awan-Ur-Rahman. (2023, August 2). *Understanding soft voting and hard voting: A comparative analysis of ensemble learning methods*. Medium.

Basher, S. A., & Sadorsky, P. (2022). *Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility?* MPRA. https://doi.org/10.2139/ssrn.4128509

Baur, D. G., & Oll, J. (2024). *The future of crypto assets and their role in the financial system*. SSRN.

Berger, T., & Koubová, J. (2024). Forecasting Bitcoin returns: Econometric time series analysis vs. machine learning. *Journal of Forecasting, 43*(7), 2904-2916. https://doi.org/10.1002/for.3165

Bistarelli, S., Claudio, A., & Santini, F. (2024). A short survey on Bitcoin price prediction. *CEUR Workshop Proceedings, 3791*.

Bouri, E., Cepni, O., Gabauer, D., & Gupta, R. (2024). Quantile connectedness in the cryptocurrency market. *Financial Innovation, 10*(1).

McFadden, F. (2025). *Competence measure enhanced ensemble learning voting schemes*. Presented at DATAWorks 2025, April 24, Alexandria, VA. https://doi.org/10.61278/itea.46.3.1007

Padmavathy, M. (2024). Behavioral finance and stock market anomalies: Exploring psychological factors influencing investment decisions. *Shanlax International Journal of Management, 11*(S1), 191-197. https://doi.org/10.34293/management.v11iiS1-Jan.7164

Phuoc, T., Hieu, P., Duy, D., & Minh, N. (2024). Applying machine learning algorithms to predict the stock price trend in the stock market - The case of Vietnam. *Humanities and Social Sciences Communications,* 11. https://doi.org/10.1057/s41599-024-02807-x

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(4), e1249. https://doi.org/10.1002/widm.1249

Schatzmann, J. E., & Haslhofer, B. (2023). Exploring investor behavior in Bitcoin: A study of the disposition effect. *Digital Finance, 5*, 581-612. https://doi.org/10.1007/s42521-023-00086-w

Shahariar, A. S. M., & Akhand, M. A. H. (2023). Implementing XGBoost machine learning model with technical indicators to predict stock price trend direction. *IEEE Xplore*.

Song, Z. (2025). Behavioral biases in the cryptocurrency market: A study on the impact of investor sentiment on price anomalies. *Journal of Applied Economics and Policy Studies, 18*(2), 36-40. https://doi.org/10.54254/2977-5701/2025.21938

Vidani, J. (2024). *Why 90% of stock market traders are in loss?* SSRN. https://doi.org/10.2139/ssrn.4849875

Appendix A

Table 1: Precision and TPM scores for each model

Model	Precision	TPM
Random Forest	0.585	48.5
CatBoost	0.575	66.1
LSTM	0.563	63.3
LGBM	0.559	53.2
XGBoost	0.557	54.9
CNN- LSTM	0.553	82.7
LogisticReg	0.551	79.5
GRU	0.528	127.1
CNN	0.525	95.3
TCN	0.523	133.6

Table 2: Precision and TPM scores for hard voting ensembles

Model	Precision	TPM
GRU + LSTM + RF + CNNLSTM	0.6157	23.3
CNN + GRU + LSTM + XGB	0.6144	27.5
CB + GRU + LGBM + CNNLSTM	0.6140	25.3
CNN + GRU + LSTM + RF	0.6138	30.0
GRU + LGBM + RF + CNNLSTM	0.6130	21.0
GRU + LGBM + LSTM + RF	0.6113	28.5
CB + GRU + LGBM + LSTM	0.6113	32.2
CB + GRU + LGBM + RF	0.6082	31.8
CB + GRU + LSTM + CNNLSTM	0.6081	28.5
CB + GRU + RF + CNNLSTM	0.6070	28.5
GRU + LR + XGB + CNNLSTM	0.6069	26.2
GRU + LGBM + XGB + CNNLSTM	0.6066	21.1
CB + CNN + GRU + LSTM + XGB + CNNLSTM	0.6066	41.1
CB + CNN + GRU + LGBM + RF + CNNLSTM	0.6063	38.3
CB + GRU + LGBM + LSTM + RF + CNNLSTM	0.6062	38.8

Table 3: Precision and TPM scores for weighted soft voting ensembles

Model	Precision	TPM
GRU + LGBM	0.6375	4.2
GRU + RF + TCN	0.6312	23.0
GRU + LGBM + TCN	0.6304	4.8
GRU + RF	0.6288	22.3
CB + GRU + LGBM + TCN	0.6250	13.2
CB + GRU + LGBM + LSTM	0.6245	24.0
CB + GRU + LGBM + LSTM + TCN	0.6240	24.5
CB + GRU + RF	0.6199	29.9
CB + GRU + LGBM	0.6194	12.6
GRU + LGBM + RF	0.6193	24.8
GRU + LGBM + LSTM	0.6184	18.3
GRU + LSTM	0.6175	14.5
CB + GRU + RF + TCN	0.6175	30.3
CB + GRU + LSTM + TCN	0.6173	20.6
GRU + LGBM + RF + TCN + CNNLSTM	0.6168	27.9

Table 4: Precision and TPM scores for weighted soft voting ensembles

Model	Precision	TPM
LGBM + CNN-LSTM	0.85714	0.4
CB + LGBM + XGB + CNNLSTM	0.83333	0.3
CB + LGBM + LSTM + CNNLSTM	0.80000	0.3
CB + LGBM	0.77778	0.5
LGBM + XGB + CNNLSTM	0.75000	0.7
CB + LGBM + XGB	0.73333	0.8
LGBM + LSTM + XGB	0.72727	0.6
CB + GRU + LGBM	0.71429	0.8
CB + GRU + LGBM + XGB + CNN-LSTM	0.70588	1.0
CB + GRU + LGBM + LR + XGB	0.70000	1.1
CB + CNN + GRU + LGBM	0.69231	0.7
CB + GRU + LGBM + LSTM + XGB	0.68750	0.9
CNN + LGBM + LSTM + RF	0.68539	5.0
CB + CNN + GRU + LGBM + XGB	0.68421	1.1
CB + CNN + LGBM + LSTM + RF + XGB	0.68235	4.8
CNN + LGBM + LSTM + RF + XGB	0.68132	5.1
CB + CNN + GRU + LGBM + LSTM + RF	0.68085	5.3
LSTM + XGB	0.67925	8.9
LGBM + LSTM + XGB	0.67901	9.1