ISSN: 2754-6705

Journal of Mathematical & Computer Applications

SCIENTIFIC Research and Community

Research Article

Open Access

From Market Noise to Signal: Machine Learning and Quantitative Alpha in Financial Markets

Mikhail Urinson

Chief Investment Officer and Data Scientist Legacy Quant | ARK Quant Crypto Miami, USA

ABSTRACT

The article examines the application of machine learning methods to the extraction of persistent alpha signals from classical technical indicators in the BTC/USD market under conditions of elevated volatility and shifting market regimes. The relevance of the study stems from the gap between the widespread use of indicators with historically established parameters and the requirements of the cryptocurrency market, which is characterized by a continuous trading cycle, a distinct microstructure, and a high level of noise. The aim of the work is to test whether a genetic algorithm can identify reproducible trading regularities that retain predictive power outside the training sample. The scientific novelty of the article is demonstrated by empirical substantiation of the thesis that classical indicators retain value under deep parametric recalibration and strict overfitting control through multistage back testing, sample splitting, and Monte Carlo stress tests. The principal findings show that alpha persists within a multidimensional parameter space. A portfolio of 11 strategies demonstrated an annual return of 26% with a maximum drawdown of 15%, and most standard indicator settings required revision. The article will be useful for researchers in quantitative finance, algorithmic traders, trading system developers, and investors studying methods for generating persistent alpha in cryptocurrency markets.

***Corresponding author**

Mikhail Urinson, Chief Investment Officer and Data Scientist Legacy Quant, ARK Quant Crypto Miami, USA.

Received: April 24, 2026; **Accepted:** April 30, 2026; **Published:** May 08, 2026

Keywords: Machine Learning, Quantitative Trading, Technical Indicators, Genetic Optimization, Back Testing, Bitcoin, Alpha Generation, Robustness Testing

Introduction

The transformation of global financial markets, catalysed by digitalization and increased computational capacity, has led to an influx of retail capital, particularly into the digital asset segment [1]. Mass access to analytical tools through popular trading platforms has given retail investors the opportunity to use methods of technical analysis that had previously remained the prerogative of professional traders [2]. This process has exposed a paradox of contemporary market microstructure. The overwhelming majority of non-institutional participants continue to use technical indicators with default parameters, such as the relative strength index RSI with a period of 14 or the moving average convergence-divergence MACD with parameters 12/26/9 [3]. These settings were derived heuristically by the pioneers of technical analysis for traditional equity and commodity markets of the previous century. They do not account for the round-the-clock trading format, extreme volatility, liquidity fragmentation, and the specific microstructure of contemporary cryptocurrency markets [4]. As a result, the application of default indicators without mathematical adaptation leads to systematic capital losses and incapacity to outperform the baseline return of a buy-and-hold strategy.

At the opposite pole of the financial ecosystem stand institutional investors and quantitative hedge funds, which have evolved from simple linear econometric models to high-performance machine learning algorithms [5]. Institutional actors actively employ genetic programming, neural networks, and ensemble learning algorithms to detect hidden nonlinear patterns [6]. The gap between these two groups generates an asymmetry of capabilities. Institutional algorithms exploit inefficiencies produced by the non-optimized actions of retail order flow. As emphasized in recent studies, the introduction of institutional discipline and a systematic approach into retail systems has become an existential imperative for market survival [7].

Within academic discourse, against the backdrop of the dominance of the efficient market hypothesis, the view remains widespread that algorithmic arbitrage and high-frequency trading have fully eliminated market inefficiencies and that alpha, excess return independent of the market beta factor, is definitively dead [8]. This study advances the opposite thesis: alpha is not dead. It is concealed within a deep multidimensional parameter space. In other words, the mathematical concepts underlying classical indicators, such as momentum, volatility, and trend, retain predictive value, yet their traditional temporal and amplitude constants are now critically obsolete.

Particular attention in contemporary quantitative finance literature is devoted to the problem of model overfitting on historical data. Researchers note that most so-called optimal investment strategies described in academic and market literature are statistical mirages generated by ignoring the correction for multiple hypothesis testing [9]. Without strict validation methods, machine learning tends to memorize historical noise rather than extract a genuine market signal.

In view of the foregoing, the main objective of this study is to systematically assess the extent to which machine learning methods, in particular genetic optimization, can extract reproducible, statistically stable alpha from classical technical indicators applied to the BTC/USD asset.

To Achieve this Objective, the Following Tasks were Formulated

1. To perform historical testing covering the generation of algorithms for the Bitcoin cryptocurrency over a long representative period.
2. To use methods of genetic evolution for multidimensional optimization of the parametric space of basic technical indicators.
3. To develop and implement a robustness verification protocol using sample splitting to eliminate the risk of overfitting.
4. To construct a final algorithmic portfolio from the most robust strategies and to verify its Out-of-Sample effectiveness across different market regimes.
5. The scientific novelty of this work lies in the empirical demonstration that adaptive parametric recalibration of classical indicators using machine learning enables overcoming the asymmetry in capabilities in financial markets. The work proposes a transparent, methodologically rigorous architecture for quantitative selection that effectively neutralizes overfitting, thereby establishing new standards for alpha generation in cryptocurrency markets.

Materials and Methodology

The methodological basis of this study is built on a hybrid approach combining systematic back testing, machine learning based on genetic algorithms, and stress testing of financial time series. The experiment's architecture was designed to ensure maximum objectivity, reproducibility, and protection against systematic sampling errors.

A comprehensive body of sources served as the foundation for the theoretical and practical components of the analysis. The empirical basis consisted of historical market data for the BTC/USD asset provided by the Bitfinex cryptocurrency exchange [10]. A daily timeframe was used, including open, high, low, and close prices for the period from January 1, 2017, to early 2025. The sample spans 2900 trading days of market phases including anomalous bullish expansion, macroeconomic shocks, prolonged downturns, structural shifts such as institutionalization of the asset, for example the approval of spot Bitcoin ETFs in 2024.

The overfitting prevention methodology was based on techniques from academic literature. Concepts of combinatorially symmetric cross-validation were implemented, demonstrating the need to use the deflated Sharpe ratio to assess the true performance of strategies after multiple testing [11, 12]. Contemporary studies on algorithmic assembling and CPCV methods were also considered [13].

Strategy generation was carried out using the author's proprietary research platform, architecturally based on the functionality of the advanced analytical environment Strategy Quant X Pro. This instrument automates the search for trading patterns without requiring a human operator to program rules, relying solely on algorithmic maximization of predefined performance criteria.

Twenty-five technical indicators were selected as elementary building blocks for strategy construction. Unlike contemporary approaches that feed raw data directly into deep neural networks, the preference for classical indicators stemmed from a need for interpretability of the resulting rules.

The indicators were conceptually grouped into five macro-categories. The first group included indicators reflecting the direction and strength of a trend. This category comprises Super Trend, Bears Power, Bulls Power, Average Directional Index with the +DI and -DI lines, Momentum, Awesome Oscillator, and Vortex.

The second group consisted of the principal entry and exit triggers. It included Linear Regression, Moving Average Convergence/Divergence, Bollinger Bands, Keltner Channel, Quantitative Qualitative Estimation, smoothed moving averages including Hull MA, as well as Ichimoku Kinko Hyo. The third category comprised oscillators intended for assessing overbought and oversold states. This group considered Commodity Channel Index, Demarket, Relative Strength Index, Laguerre RSI, Schaff Trend Cycle, and Williams Percent Range.

The fourth macro-category covered metrics associated with risk management and volatility assessment. It included Average True Range, Ulcer Index, and Standard Deviation. The fifth group consisted of pure trend filters. It included the Kaufman Efficiency Ratio, Aroon, and Parabolic SAR.

The alpha search process was based on the method of genetic evolution. Unlike random generation, which conducts an extensive search of the space without regard to accumulated experience, the genetic algorithm functions according to the principles of biological natural selection. The algorithm initiates an initial population of random solutions, evaluates their fitness in terms of risk and return indicators, and then selects the fittest strategies for crossover, the recombination of trading rules, and mutation, the modification of individual parameters. Optimization was limited to 100 generations. To prevent the algorithm from becoming trapped in local optima, an island model was used. Evolution proceeded in parallel across four isolated subgroups, each with 100 strategies and periodic migration between islands, exchanging 1–5% of the best individuals.

The environment configuration included 32 key parameters. The strategies used both symmetric and asymmetric logic for entering Long and Short positions. To prevent combinatorial explosion and excessive model complexity, the number of mandatory entry and exit conditions was limited to 1-4 to implement fuzzy logic. The indicator calibration procedure became the most important step. Indicator parameters were constrained within an effective range equal to the historical mean plus or minus the standard deviation. Forced risk limitation was also applied. Stop Loss and Profit Target were set in dynamic ranges linked to volatility, ATR from 1.5 to 5, which constitutes best practice in capital management. The fitness function used for algorithm ranking was the Target Threshold criterion, with a threshold of> 4.

To avoid statistical overfitting and the hindsight bias, the quotation dataset was rigidly split into 40% / 30% / 30% proportions, fully consistent with advanced machine learning standards in multilayer investment analysis.

The In-Sample Training period comprised 40% of the sample and covered the years 2017-2020. These data were used by the genetic algorithm as the training environment for forming trading rules.

The In-Sample Validation period comprised 30% of the sample and covered the years 2020-2022. It represented an independent sample intended to assess algorithm convergence and reduce the risk of overfitting on the training set. When the population fitness function began to stagnate at the ISV stage, the evolutionary process on the respective island was restarted.

The Out-of-Sample period also comprised 30% of the sample and covered the years 2022-2024+. This time interval remained completely inaccessible to the generator and was used as the final performance evaluation criterion. This period included one of the most difficult market cycles, including the 2022 bear market.

After more than 10 million strategies had been generated, the algorithmic filter eliminated the overwhelming majority. Only those strategies that surpassed the basic threshold values, Net Profit > $10,000, Return/DD > 4, Profit Factor > 1.3, number of trades > 100, Win Rate > 30%, were retained in the primary database. There were more than 10,000 such candidates.

A robust testing protocol was developed and implemented through Monte Carlo simulations. Each of the 10,000 strategies underwent two types of stress tests, enabling assessment of algorithm stability under random distortions and adverse execution conditions.

The first type was Monte Carlo Trades Manipulation. In this simulation, the historical trade sequence was randomly shuffled using the Randomize Trades Order procedure. In addition, random trades were artificially removed from the sequence using the Random Skip Trades mechanism, which modelled a loss of exchange connectivity or a liquidity deficit. Under such distortions, the algorithm was required to maintain a positive mathematical expectation even when profitable outliers were omitted.

The second type included Monte Carlo Retest Methods, designated as What-If Scenarios. This was a more resource-intensive test, in which the strategy was fully recalculated with random noise added to the historical data via Randomize History Data. At the same time, a shift in the starting bar and microscopic changes to strategy parameters were applied via Randomize Strategy Parameters. This approach enabled testing the strategy's sensitivity to small variations in source information and model configuration.

Only those algorithms whose equity curves-maintained stability at the 95% confidence level were transferred to the passed status. As a result, an experimental pool of 100 strategies was formed from more than 10 million generated concepts, designated as Top-100 deployed. From this set, after additional microstructure analysis, the final portfolio Top-11 currently in use was selected and launched into live trading.

Results and Discussion

The aggregated results of the 11 best strategies that passed all stages of robust filtering demonstrate outstanding risk-adjusted efficiency. As presented in Table 1, the strategy portfolio delivers absolute returns and institutional-grade risk metrics.

Table 1: Summary of the Top 11 Algorithmic Strategies Portfolio

Key Metric	Baseline Benchmark, Target	Actual Result of the Top-11 Portfolio
Average Net Profit	—	$21,502
Average Annual Return	25.0%	26.0%
Win Rate	60.0%	64.0%
Profit Factor	2.0	2.0
Win/Loss Ratio	2.0	1.9
Return/Drawdown Ratio	10.0	10.7
Maximum Drawdown	< 20.0%	15.0%

The portfolio's average annual return was 26%, which meets the target benchmark. The most important indicators of robustness are a Win Rate of 64% and a Profit Factor of 2.0. This means the strategies' expected value is firmly in the profitable zone. Maximum drawdown is tightly controlled and averages only 15% of capital, yielding a Return/DD of 10.7. These metrics confirm that the portfolio is not exposed to the catastrophic black swan risks characteristic of unsystematic cryptocurrency trading. Synergistic diversification is observed within the portfolio. Momentum models could return nearly 30%, and several RSI/QQE variants smoothed the equity curves, reducing maximum drawdowns to 9-13%.

Strategy #1 is, however, a sort of benchmark to machine learning alpha extraction, in the sense that a model outperforms the customary equity markets. The model makes 1,618 trades through the analysed period. With a fixed position size, the cumulative profit would have been $48,034 for a 480% return. The average annual growth rate was recorded at 28.56%. Under an alternative Return/Year calculation methodology without reinvestment, the indicator is estimated at 44%.

The exceptional nature of the algorithm is revealed through its risk metrics. The Sharpe ratio reached 3.53, the Profit Factor amounted to 5.72, and the Return/DD ratio set an absolute record of 206. The maximum drawdown over more than 7 years amounted to only 0.47%. With a successful trade percentage of 55.19%, the strategy demonstrates the classical algorithmic pattern of cutting losses and expanding profits.

The secret of such phenomenal stability lies in the algorithm’s use of a machine-generated fuzzy logic concept. The source code for Strategy #1 shows that trading decisions are made by weighting a set of conditions. Thus, the signal for opening a long position is activated only when at least 46% of logical conditions are satisfied, for example, 2 out of 5 complex blocks including non-standard indicator states of Keltner Channel, Bollinger Bands, MACD, and Super Trend. The position-closing signal is even more stringent. It requires confirmation of 77% of conditions (5 out of 6), integrating analysis of Vortex, QQE, Momentum, and reversals in the Awesome Oscillator. Machine learning enabled the construction of a multilayer filter that filters out market noise and reacts only to macro-structural shifts in the probability density of prices.

A deep understanding of the nature of algorithmic alpha requires analysis of model behavior across different market regimes. The property of algorithms to dynamically adjust to macroeconomic volatility is a key aspect of constructing portfolios resistant to paradigm shifts. The visual object below presents the aggregated annual return of the constructed portfolio. The dynamics of the total return of the Top-11 algorithmic strategies portfolio by calendar year are shown in Figure 1.

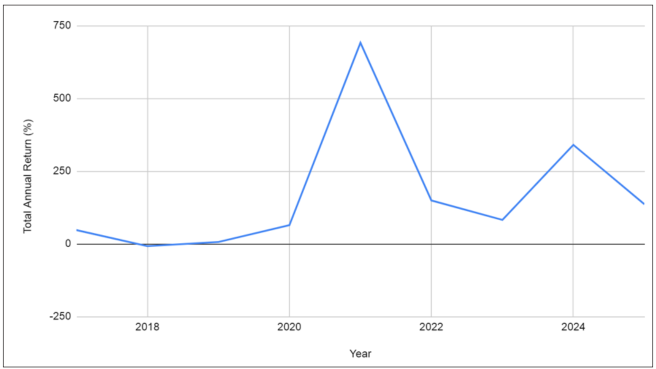

Figure 1: Total Return Dynamics of the Top 11 Algorithmic Strategies Portfolio by Calendar Year

The data in Figure 1 illustrate the system's high adaptability. During the macroeconomic euphoric expansion of 2021, the bull Supercycle, the portfolio of algorithms demonstrated a phenomenal capacity to capitalize on the trend, delivering a cumulative Total Return of +691%. Some individual strategies within the portfolio showed triple-digit returns, for example +109%, +101%, and +86%.

The true test for any quantitative model is the bear market. In 2018, when Bitcoin capitalization collapsed by more than 70%, the algorithmic portfolio showed a cumulative loss of only -8%. This was the only negative year over the entire study period. Against the backdrop of the global collapse of the crypto industry in 2022, the implosion of the Terra/Luna ecosystem and the FTX exchange, the portfolio preserved capital and extracted gains from extreme volatility, generating a return of 149%. Recent dynamics also confirm the models' viability. Amid the institutionalization of Bitcoin in 2024, cumulative return amounted to 340%, and by the beginning of 2025, it had already reached 136%.

Technical indicators are still relevant and only their constants have become obsolete. The analysis showed that 80% of the parameters of classical indicators were altered and optimized by the genetic evolution algorithm for use in cryptocurrency work. Table 2 is presented to illustrate the scale of the parametric changes.

Table 2: Comparative Analysis of Classical and ML-Optimized Parameters of Technical Indicators

Indicator and Parameter Name	Classical, Default Value	Value Range After ML Optimization, Strategies 1–11	Interpretation of Changes
MACD, Fast / Slow / Signal	12 / 26 / 9	(7 / 26 / 8), (4 / 41 / 4), (8 / 52 / 9)	Significant acceleration of the fast line to react to sharp jumps in crypto market microstructural volatility
RSI, Period	14	20	Increase in the period to filter out high-frequency intraday noise and avoid false entries on false breakouts
ADX, Lower Period	14	30	Extension of the data collection horizon for more accurate detection of the start of a true macro-trend
Bollinger Bands, Bar Opens Period	20	147	Radical transformation from a short-term channel oscillator into a macro-indicator of global regime change
Standard Deviation, Higher Period	20	185	Similar to Bollinger Bands, adapted for analyzing structural seasonality

The cause of the total inefficiency of classical parameters lies in market physics. Settings such as 14 days for RSI or 20 days for Bollinger Bands were selected heuristically in an era when markets traded 5 days per week, with 14 days equaling 3 working weeks and 20 days equaling 1 trading month. Applying these constants to Bitcoin, which trades continuously 24/7/365, creates a phase shift in mathematical expectations. The genetic algorithm, unconstrained by human heuristics, successfully recalibrated this frequency disproportion.

A vulnerability in quantitative finance is the illusion of predictability generated by overfitting to historical data. The multiple testing effect means that when millions of strategies are tested on a limited data set, algorithms with an ideal equity curve will inevitably appear purely by statistical chance [14]. The robustness-testing architecture implemented in the study, shown in Figure 2, enabled the mathematical proof of the presence of systemic alpha.

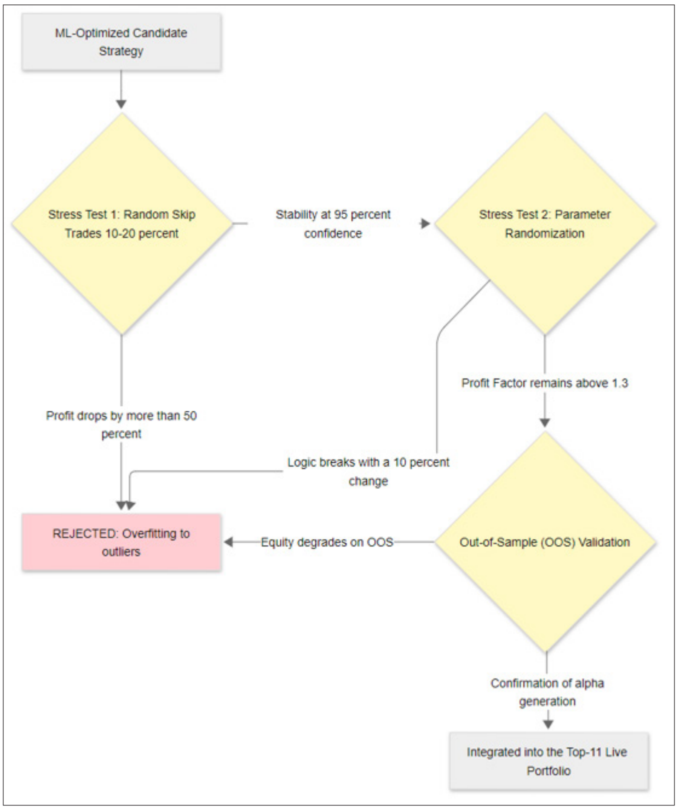

Figure 2: Robustness Test Architecture for Avoiding Statistical Overfitting

Analysis of Monte Carlo results with trade manipulation showed that random exclusion of 10% of trades, which emulates real market conditions such as liquidity problems, network latency, or exchange API failures, leads only to the mathematically expected reduction in profit and does not collapse the strategy's equity. At the 95% confidence level, the algorithms continued to demonstrate high returns and acceptable drawdowns. The survival of the algorithms in the Out-of-Sample zone, which comprised 30% of the entire data set and included the turbulence of 2022–2024, provides final confirmation that genetic optimization extracted a true market pattern.

The results obtained open new horizons for understanding the nature of the market signal. Machine learning, in particular, evolutionary algorithms, addresses the task of alpha extraction efficiently for several reasons. First, the parameter space of 25 indicators has limited dimensionality, yet remains too complex and nonlinear for direct brute-force search. Second, using a fitness function based on the Return/DD ratio forces the machine to search for parameter values that yield a high, smooth, and consistent equity curve. The machine finds quiet harbors of parameters that remain stable under minor noise perturbations in quotations.
From a practical standpoint, this study carries large-scale implications. For retail investors, it provides a mathematical justification for abandoning the default settings of trading terminals. For institutional participants, it confirms that integrating algorithmic discipline and multilayer decision-making frameworks can create strategies resistant to shifts in market paradigms. Within academic discourse, this work serves as a transparent methodological precedent for structuring a hypothesis-testing pipeline to neutralize back test overfitting.

At the same time, the study has certain limitations that must be taken into account. Testing was conducted on the most liquid crypto asset, Bitcoin, whose microstructure differs substantially from that of low-liquidity altcoins, where slippage and market impact in managing large capital can materially distort Profit Factor values. In addition, the specificity of the selected fitness function forms a certain style bias in the algorithms. The algorithms become highly risk-averse, leading to premature exits during aggressive bull rallies. Finally, the absence of dynamic accounting for transaction costs under changing liquidity conditions leaves room for subsequent conceptual development of the proposed framework.

Conclusion

The study provides comprehensive empirical evidence that machine learning methods can transform chaotic market noise into a structured alpha signal. Using ultra-large-scale back testing of more than 10 million algorithmic combinations on historical Bitcoin data from 2017–2025, it was scientifically confirmed that technical indicators retain predictive validity. The unprofitability of traditional approaches stems from the fatal obsolescence of parametric constants that do not reflect the physics of the modern, round-the-clock cryptocurrency market.

Solving the stated research tasks enabled the generation of a composite portfolio of 11 strategies via genetic evolution. The portfolio results demonstrate outstanding institutional efficiency: the average annual return reached 26%, the Win Rate rose to 64%, and maximum drawdown remained strictly controlled at 15%. A detailed analysis of the leading strategy showed that applying machine-generated fuzzy logic, combining multiple indicators with optimized parameters, ensures reliable capital protection during market collapses.

The central thesis of the work has been fully confirmed: alpha is not dead; it is concealed in deep parameter spaces. The analysis showed that the ML algorithm substantially adjusted most of the classical indicator parameters to better reflect actual volatility, for example by increasing the Bollinger Bands period and making the MACD more responsive. Passage through Monte Carlo simulations with trade manipulation and strict Out-of-Sample control in a 40/30/30 proportion mathematically proves the absence of overfitting.

The practical significance of the work lies in creating a verifiable algorithmic framework that neutralizes the technological gap between institutional and retail investors. Further research directions include extending the optimization methodology to cross-asset portfolios, integrating dynamic transaction cost models, and applying causal machine learning concepts to ensure even greater adaptability of trading systems in real time.

References

1. Havakhor T, Rahman MS, Zhang T, Zhu C (2024) Tech-enabled financial data access, retail investors, and gambling-like behavior in the stock market. Manag Sci 71: 1-25.
2. Chio PT (2022) A comparative study of the MACD-base trading strategies: evidence from the US stock market https://arxiv.org/abs/2206.12282.
3. Kang BK (2021) Improving MACD technical analysis by optimizing parameters and modifying trading rules: evidence from the Japanese Nikkei 225 futures market. J Risk Finance Manag https://www.mdpi.com/1911-8074/14/1/37.
4. Easley D, O Hara M, Yang S, Zhang Z (2026) Microstructure and market dynamics in crypto markets. J Finance Mark https://www.sciencedirect.com/science/article/abs/pii/S1386418126000261.

5. Shi C (2025) From econometrics to machine learning: transforming empirical asset pricing. J Econ Surv 40: 28-48.
6. Christodoulaki E, Kampouridis M, Kyropoulou M (2025) A novel strongly-typed genetic programming algorithm for combining sentiment and technical analysis for algorithmic trading. Knowl Based Syst https://www.sciencedirect.com/science/article/pii/S0950705125001017.
7. Das AR, Chakraborty A, Goyal N, Roy P (2025) Financial technology in transition: trends development and ways forward. J Econ Surv https://onlinelibrary.wiley.com/doi/10.1111/joes.70029?utm_source=researchgate.net&utm_medium=article.
8. Ghatak S, Khaledian A, Parvini N, Khaledian N (2025) Increase alpha: performance and risk of an AI-driven trading framework https://arxiv.org/abs/2509.16707.
9. Anghel DG (2021) Data snooping bias in tests of the relative performance of multiple forecasting models. J Bank Finance https://www.sciencedirect.com/science/article/abs/pii/S03784266210.
10. (2026) Bitfinex Candles https://docs.bitfinex.com/reference/rest-public-candles.
11. Arian H, Mobarekeh DN, Seco L (2024) Back test overfitting in the machine learning era: a comparison of out-of-sample testing methods in a synthetic controlled environment. Knowl Based Syst https://www.sciencedirect.com/science/article/abs/pii/S0950705124011110.
12. Pagliaro A (2026) Regime-aware Light GBM for stock market forecasting: a validated walk-forward framework with statistical rigor and explainable AI analysis. Electronics https://www.mdpi.com/2079-9292/15/6/1334.
13. Sun Y, Qu Z, Zhang T, Li X (2025) Adaptive ensemble learning for financial time-series forecasting: a hypernetwork-enhanced reservoir computing framework with multi-scale temporal Modeling https://www.mdpi.com/2075-1680/14/8/597.
14. Andrews JP (2025) A replication study of Zhang (2025): why news sentiment fails to predict asset returns https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6426038.