Applied Decision Analytics Volume 2, Issue 1 (2026) 124-140

Applied Decision Analytics

Journal homepage: www.ada-journal.org
ISSN: 3104-2945

Interpretable Extreme Gradient Boosting Models for Multi-Class Stock Movement Prediction Using Technical Indicators

Bharatendra Rai[1,*]

[1] Department of Decision and Information Sciences, Charlton College of Business, University of Massachusetts – Dartmouth, USA

ARTICLE INFO

Article history:
Received 26 August 2025
Received in revised form 27 October 2025
Accepted 15 November 2025
Available online 26 December 2025

Keywords:
Extreme gradient boosting, Stock market prediction, Technical indicators, Hyperparameter tuning

ABSTRACT

Accurately forecasting short-horizon stock price movements remains a challenging problem in financial markets, as classical financial theory suggests that prices in liquid markets quickly incorporate available information. Nevertheless, empirical evidence continues to document short-term patterns related to momentum, volatility, and regime shifts, motivating the use of technical indicators within modern machine learning frameworks. While recent studies report promising predictive performance, many emphasize aggregate accuracy, apply limited hyperparameter tuning, or provide little interpretability, leaving uncertainty about the robustness and economic meaning of their results. The purpose of this study is to evaluate whether Extreme Gradient Boosting models built exclusively on technical indicators can reliably and transparently classify short-term stock movements into down, neutral, and up regimes. Using daily data for three large-cap U.S. equities—Tesla, Apple, and Nvidia—we construct a feature set of 36 price- and volume-based technical indicators. A two-stage hyperparameter optimization strategy combining adaptive cross-validation with refined grid search is employed, followed by strict out-of-sample evaluation using a chronological train–test split. Model interpretability is examined through global feature importance and local, instance-level explanations. The results show consistently higher sensitivity for directional movements than for neutral regimes across all assets, with momentum-based indicators, particularly Rate of Change, emerging as the dominant predictors. Local explanations reveal that strong alignment among momentum and oscillator signals underpins confident directional predictions, whereas mixed signals lead to neutral classifications. Overall, the findings demonstrate that well-tuned and interpretable models can extract economically meaningful structure from technical indicators, while also clarifying the limitations of predicting low-signal market regimes.

1. Introduction

Forecasting short-horizon stock price movements has long been a central challenge in financial economics and quantitative finance. Classical financial theory, particularly the Efficient Market Hypothesis (EMH), posits that asset prices in liquid markets rapidly incorporate available information,

* *Corresponding author.*
E-mail address: brai@umassd.edu

https://doi.org/10.66972/ada21202618

leaving little room for systematic predictability based solely on historical prices or volumes. Despite this theoretical benchmark, a substantial empirical literature has documented persistent patterns such as momentum, mean reversion, and volatility clustering, especially over short to medium horizons. These regularities have motivated continued interest in technical analysis as a mechanism for extracting informative signals from price and volume dynamics.

Technical indicators including moving averages, momentum oscillators, volatility bands, and volume-based measures, represent nonlinear transformations of historical market data designed to capture trend strength, regime shifts, and overbought or oversold conditions. While early applications of technical analysis were often heuristic, more recent research has embedded these indicators within formal statistical and machine learning frameworks. Advances in computational power and algorithmic design have enabled models capable of capturing nonlinear interactions and threshold effects among indicators, yielding improved predictive performance relative to linear benchmarks in certain market settings.

The adoption of machine learning methods has further expanded the scope of empirical asset return prediction. Algorithms such as support vector machines, random forests, gradient boosting machines, and deep neural networks have been widely applied to equity, index, and derivative markets. Among these, tree-based ensemble methods have gained particular prominence due to their ability to model nonlinearities, handle heterogeneous feature sets, and deliver competitive accuracy with relatively modest data requirements. Extreme Gradient Boosting (XGBoost), in particular, extends classical boosting frameworks by incorporating regularization, efficient tree construction, and scalable optimization, making it well suited for high-dimensional technical indicator inputs.

Despite these advances, several limitations remain in the existing literature. First, many studies emphasize aggregate predictive accuracy while giving limited attention to class-specific performance, especially in multi-class settings that distinguish between upward, downward, and neutral price movements. Second, hyperparameter tuning is often treated superficially, with coarse grids or ad hoc settings that obscure the role of regularization and learning dynamics in driving results. Third, although interpretability has become an increasingly important concern in financial machine learning, relatively few studies combine strong predictive models with systematic, instance-level explanations that clarify how technical indicators jointly influence individual predictions. As a result, it remains unclear whether reported performance gains reflect economically meaningful structure or model-specific artifacts.

This study addresses these gaps by providing a systematic, reproducible, and interpretable evaluation of XGBoost-based multi-class stock movement classification using a comprehensive set of technical indicators. Focusing on three large-cap, highly traded U.S. equities—Tesla, Apple, and Nvidia—we go beyond aggregate accuracy by examining class-wise sensitivity (recall) for down, neutral, and up movements, thereby shedding light on asymmetries in predictive performance across market regimes. We further contribute by implementing a two-stage hyperparameter optimization strategy, combining adaptive cross-validation with refined grid search, to ensure robust model selection without excessive data snooping. Finally, we integrate global feature importance analysis with local, instance-level explanations, explicitly linking indicator-level decision logic to observed class-wise performance.

The objectives of this research are threefold:

i. to benchmark the out-of-sample performance of XGBoost models built solely on technical indicators for multi-class stock movement prediction;
ii. to analyze how different indicators contribute to predictions at both the global and local levels; and

iii. to connect interpretability results to class-specific sensitivity outcomes, thereby improving understanding of when and why machine learning models succeed or struggle in distinguishing directional versus neutral market regimes.

By unifying rigorous evaluation, transparent tuning, and interpretable explanations, this study contributes to the growing literature on explainable financial machine learning and provides practical insights for researchers and practitioners seeking reliable, interpretable models for short-horizon market prediction.

2. Literature Review

Empirical research on stock return predictability spans several decades and documents both limits and pockets of exploitable structure in financial time series. Early works grounded in the Efficient Market Hypothesis (EMH) argued that, in liquid markets, past prices and volumes should contain little information about future returns beyond what is implied by risk exposures. Nonetheless, a large body of evidence has accumulated showing short- to medium-horizon serial dependence, momentum, and mean reversion effects, particularly when returns are conditioned on technical or macroeconomic state variables [1-3]. Technical analysis, once regarded as ad hoc, has increasingly been framed in terms of nonlinear transformations of prices and volumes that can serve as inputs to statistical and machine learning models, with recent studies demonstrating the effectiveness of indicators such as moving averages, momentum, volatility measures, RSI, and MACD in generating superior trading signals and cumulative returns when combined with modern algorithms [4-7].

With the rise of machine learning, researchers began to replace linear factor models with more flexible classifiers and regressors such as support vector machines, random forests, gradient boosting machines, and deep neural networks [8]. Studies have reported that tree-based ensemble methods often strike a useful balance between predictive accuracy, robustness, and interpretability, particularly when working with technical indicators that exhibit nonlinear interactions and threshold effects [9-11]. Gradient boosting, in particular, has been shown to capture complex conditional relationships between indicator configurations and subsequent price moves, while providing calibrated probabilities that can be fed into trading and risk-management frameworks [12].

Extreme Gradient Boosting (XGBoost) extends classical gradient boosting by incorporating advanced regularization, efficient tree construction, and native handling of sparse and high-dimensional data. In financial applications, XGBoost has been used for equity and index direction prediction, credit risk modelling, option pricing approximations, and portfolio risk forecasting, with enhancements like multi-objective optimization improving risk-adjusted returns [13-14]. Several comparative studies find that XGBoost and other tree-based ensembles frequently outperform or match deep learning architectures on tabular financial datasets, especially when the sample size is moderate and the feature set consists of heterogeneous hand-crafted indicators rather than raw images or text [15-17]. At the same time, concerns about overfitting, instability across regimes, and data snooping remain central, highlighting the importance of careful evaluation protocols such as independent train-test splits, multiple performance metrics, and external validation [18].

Against this backdrop, the present study contributes by providing a focused, systematic benchmark of XGBoost models built purely on technical indicators for three large-cap, highly traded equities. Rather than proposing a novel architecture, we emphasize rigorous hyperparameter search, transparent reporting of cross-validation and out-of-sample results, and an interpretable analysis of indicator relevance via feature importance. This aligns with recent calls in the literature for more reproducible, methodical comparisons instead of isolated case studies with opaque modelling choices [19-22].

3. Data and Methodology

3.1 Data Description

The empirical analysis is conducted on daily data for three U.S.-listed equities: NVIDIA (NVDA), Apple (AAPL), and Tesla (TSLA). For each security, we construct a panel of technical indicators based on historical prices and volumes. The indicators include, but are not limited to, simple and exponential moving averages (SMA, EMA) over multiple horizons, the Average Directional Index (ADX), Aroon oscillator, Trader's Dynamic Index (TDI), Donchian channels, the Vertical Horizontal Filter (VHF), Commodity Channel Index (CCI), Money Flow Index (MFI), On-Balance Volume (OBV), Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), stochastic oscillators, Chande Momentum Oscillator (CMO), rate-of-change (ROC), Know Sure Thing (KST), TRIX, and several band- and volatility-based measures. All 36 features are computed in a rolling fashion using only information available up to time ‘t’ to avoid look-ahead bias. For more details on the data see [4].

The dependent variable ‘y’ is defined as a discrete, three-class encoding of the next-period price movement as up, down, or neutral. Eq. (1) illustrates classification into these three categories.

$$y_t = \begin{cases} up, if\ 100 \times \frac{C_t - M_n}{M_n} > T \\ down, if\ 100 \times \frac{C_t - M_n}{M_n} < -T \\ neutral, if\ -T \le 100 \times \frac{C_t - M_n}{M_n} \le T \end{cases} \tag{1}$$

Where, y_t is one-day ahead classification of stock price in time period t, M_n is the moving average of n days before t, C_t is the closing stock price value on day t, and T is a percentage threshold (e.g., 0.5%). In this study we use n = 5, which means one-day ahead closing stock price is compared to average of closing stock price of previous five days. The threshold values used for Apple, Nvidia and Tesla in this study are 1%, 2% and 2% respectively. This study uses a quantitative financial modelling framework available from R programming package ‘quantmod’ to obtain data that is sourced from Yahoo. This data is for Apple, Nvidia and Tesla involving time period from Jan 1st, 2018, to June 30th, 2025.

3.2 Train–Test Protocol

Because financial time series are temporally ordered and potentially non-stationary, standard random shuffles are inappropriate. Instead, we adopt a chronological split: for each asset, the first 1,200 observations form the training set, and all subsequent observations constitute the test set. The test set thus represents a genuinely held-out future period. Within the training window, hyperparameters are tuned via three-fold cross-validation that respects the original ordering within each fold while still providing multiple train/validation splits. This approach balances the need for robust performance estimates against the risk of overfitting to a single split.

No explicit feature scaling or normalization is performed because tree-based ensembles such as XGBoost are invariant to monotonic transformations of the input features. All explanatory variables are treated as continuous.

3.3 Extreme Gradient Boosting Specification

The extreme gradient boosting classifier was parameterized to balance predictive flexibility with strong regularization and stable learning behavior. Model complexity was controlled through limits on tree depth, restricting the order of feature interactions while preserving the ability to capture nonlinear relationships. Learning dynamics were governed by the learning rate and the number of boosting iterations, enabling gradual error correction and stable convergence of the ensemble. Structural regularization was imposed by requiring a minimum reduction in the objective function for additional splits and by enforcing minimum node weights, both of which discourage weak or noise-driven partitions. Further variance reduction was achieved through stochastic subsampling of training observations and predictor variables at each boosting iteration, increasing diversity among trees and improving generalization performance. Together, these hyperparameter choices produce a conservative yet expressive boosted tree model that is well suited for noisy or high-dimensional classification settings.

Key hyperparameters govern the complexity and regularization of the model. The number of trees (nrounds) and maximum depth (max_depth) control the capacity of the ensemble where larger values can capture more intricate patterns but raise the risk of overfitting. The learning rate (eta) scales the contribution of each new tree and trades off convergence speed against stability. Row subsample (subsample) and column subsample (colsample_bytree) define the fraction of training rows and features, respectively, that are randomly sampled when building each tree, acting as implicit regularizers and promoting diversity among trees. The minimum sum of instance weights (min_child_weight) required in a leaf node constrains the minimum effective sample size, and gamma sets the minimum reduction in the objective function required to make an additional split, both discouraging marginally beneficial splits.

3.4 Hyperparameter Search and Model Selection

To identify well-performing configurations without exhaustively enumerating the high-dimensional hyperparameter space, we apply search in two phases as shown in Figure 1.

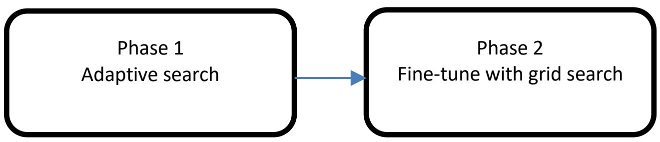

Fig. 1. Two phases for hyperparameter tuning

In phase-1, to improve computational efficiency during hyperparameter optimization while preserving statistical validity, we employed an adaptive cross-validation procedure. Model evaluation followed an adaptive resampling scheme in which candidate configurations were iteratively assessed and statistically inferior models were eliminated early. Each candidate model was first evaluated on a minimum of two resampling iterations. After this initial phase, model performance was compared using generalized least squares, which accounts for correlation among repeated resamples. At each adaptive step, models whose performance was significantly worse than the leading configuration were discarded based on a two-sided hypothesis test at $\alpha = 0.05$. This process dynamically reduced the candidate set, concentrating resampling effort on competitive models.

Following adaptive elimination, the remaining configurations were fully evaluated using the complete resampling design, ensuring unbiased and comparable performance estimates for final model selection. Parallel execution was enabled to further reduce training time. This adaptive resampling strategy balances computational efficiency and inferential rigor, avoiding exhaustive evaluation of poorly performing configurations while maintaining statistically principled model comparisons during random hyperparameter search.

In phase-2, following adaptive cross-validation, we conducted a focused expanded grid search centered on the best-performing hyperparameter configuration identified during the adaptive phase. The purpose of this second-stage search was to refine the local neighborhood of the adaptive optimum and reduce the risk that early model elimination might exclude near-optimal configurations. Specifically, hyperparameter ranges were narrowed and densified around the selected configuration, and all candidate models within this refined grid were evaluated using the full resampling design. Unlike the adaptive phase, no early stopping or model elimination was applied, ensuring that each configuration received an identical number of resamples and enabling unbiased fine-grained comparison.

This two-stage optimization strategy combines the global efficiency of adaptive search with the local precision of exhaustive evaluation, yielding robust hyperparameter selection while maintaining tractable computational cost. By separating broad exploration from targeted exploitation, the approach improves stability and reproducibility of the final model.

4. Hyperparameter Tuning Results

Hyperparameter tuning was conducted using the R programming environment. All computations were performed on an Apple Mac Studio equipped with an M2 Ultra processor and 128 GB of memory. The complete hyperparameter optimization process required approximately 21 hours of computation time. The resulting tuning outcomes for the Extreme Gradient Boosting models applied to Tesla, Apple, and Nvidia are summarized below.

4.1 Tesla Hyperparameter Search

For each hyperparameter, Table 1 provides adaptive search range from phase-1 and final results from phase-2 for Tesla.

Table 1
Phase-1 and phase-2 searches for Tesla

Hyperparameter	Adaptive search range (phase-1)	Fine-tune result (phase-2)
nrounds	5 to 989	405
max_depth	1 to 10	5
eta	0.0039 to 0.5997	0.121
gamma	0.0354 to 9.9927	7.75
colsample_bytree	0.3020 to 0.6999	0.5
min_child_weight	0 to 20	5
subsample	0.2500 to 0.9982	0.3

As seen from Table 1, for Tesla the ensemble consisted of 405 boosting iterations (nrounds = 405), allowing the model to iteratively refine predictions while maintaining stable convergence. Individual trees were restricted to a maximum depth of five (max_depth = 5), limiting interaction complexity and reducing the risk of overfitting. A learning rate of 0.121 (eta = 0.121) was used to moderate the contribution of each tree, balancing convergence speed with generalization. Tree growth was made deliberately conservative by requiring a minimum loss reduction of 7.75 (gamma = 7.75) for additional splits and enforcing a minimum child weight of five (min_child_weight = 5), both of which discourage weak or noise-driven partitions. Further regularization was introduced through stochastic subsampling, with 50% of predictors sampled per tree (colsample_bytree = 0.5) and 30% of training observations sampled per boosting iteration (subsample = 0.3), increasing ensemble diversity and reducing variance. Collectively, these settings define a strongly regularized boosted tree model that balances expressive power with robust generalization.

4.2 Apple Hyperparameter Search

For each hyperparameter, Table 2 provides adaptive search range from phase-1 and final results from phase-2 for Apple.

Table 2
Phase-1 and phase-2 searches for Apple

Hyperparameter	Adaptive search range (phase-1)	Fine-tune result (phase-2)
nrounds	5 to 989	135
max_depth	1 to 10	8
eta	0.0039 to 0.5997	0.52
gamma	0.0354 to 9.9927	5.75
colsample_bytree	0.3020 to 0.6999	0.55
min_child_weight	0 to 20	19
subsample	0.2500 to 0.9982	0.8

The final XGBoost classifier was trained using 135 boosting iterations (nrounds = 135), resulting in a relatively compact ensemble in which fewer trees were required to reach convergence. Individual trees were allowed to grow to a maximum depth of eight (max_depth = 8), enabling the model to capture higher-order nonlinear interactions among predictors. To support rapid learning with a smaller number of trees, a high learning rate (eta = 0.52) was employed, increasing the contribution of each tree and accelerating convergence, albeit with a greater risk of overfitting. This risk was mitigated through multiple regularization mechanisms. Tree growth was constrained by a minimum loss reduction threshold of 5.75 (gamma = 5.75), preventing weak or marginal splits, and a large minimum child weight (min_child_weight = 19), which discourages splits that isolate small or noisy subsets of observations. Additional variance reduction was achieved through stochastic subsampling, with 55% of predictors sampled per tree (colsample_bytree = 0.55) and 80% of training instances sampled per boosting iteration (subsample = 0.8), introducing randomness that improves generalization. Overall, this configuration reflects a model that emphasizes expressive tree structures and fast learning, counterbalanced by strong structural and sampling-based regularization to maintain robustness.

4.3 Nvidia Hyperparameter Search

For each hyperparameter, Table 3 provides adaptive search range from phase-1 and final results from phase-2 for Nvidia.

Table 3
Phase-1 and phase-2 searches for Nvidia

Hyperparameter	Adaptive search range (phase-1)	Fine-tune result (phase-2)
nrounds	5 to 989	560
max_depth	1 to 10	3
eta	0.0039 to 0.5997	0.009
gamma	0.0354 to 9.9927	1.7
colsample_bytree	0.3020 to 0.6999	0.37
min_child_weight	0 to 20	13
subsample	0.2500 to 0.9982	0.3

The final XGBoost classifier was trained using 560 boosting iterations (nrounds = 560), resulting in a large ensemble designed to learn complex patterns through gradual refinement rather than aggressive tree updates. Individual trees were constrained to a maximum depth of three (max_depth = 3), producing shallow learners that limit interaction complexity and reduce variance. To support this conservative structure, a very small learning rate (eta = 0.009) was employed, ensuring that each tree contributed only marginally to the ensemble and promoting stable convergence with strong generalization. Tree growth was further regularized by requiring a minimum loss reduction of 1.7 (gamma = 1.7) for additional splits and enforcing a minimum child weight of 13 (min_child_weight = 13), both of which discourage weak, noise-driven partitions. Additional regularization was introduced through stochastic subsampling, with 37% of predictors sampled per tree (colsample_bytree = 0.37) and 30% of training observations sampled per boosting iteration (subsample = 0.3), increasing ensemble diversity and reducing overfitting. Collectively, this configuration emphasizes slow, stable learning with strong regularization, favoring robustness and generalization over rapid convergence.

5. Out-of-Sample Performance

Out-of-sample performance for the extreme gradient boosting classification models is assessed using the test data. Table 4 reports class-specific sensitivity (recall) for a three-class classification—down, neutral, and up—evaluated separately for each company. Sensitivity measures the model's ability to correctly identify observations belonging to a given class, with higher values indicating better detection of true positives.

Table 4
Sensitivity for down, neutral and up categories

Company	Sensitivity (down)	Sensitivity (neutral)	Sensitivity (up)
TESLA	0.7018	0.4335	0.7394
APPLE	0.7320	0.3102	0.6875
NVIDIA	0.6544	0.4581	0.7005

As observed from Table 4, for Tesla, the model demonstrates relatively strong performance in identifying directional movements, with high sensitivity for both down (0.7018) and up (0.7394) classes. However, performance for the neutral class is notably lower (0.4335). This suggests the model is more effective at capturing pronounced price movements than periods of stability.

A similar pattern is observed for Apple, where sensitivities for down (0.7320) and up (0.6875) movements exceed that of the neutral class (0.3102). The particularly low neutral sensitivity indicates difficulty in distinguishing sideways or low-volatility behavior relative to clearer directional changes.

For Nvidia, sensitivities are more balanced across classes, though directional classes again outperform the neutral class. The model achieves moderate sensitivity for down (0.6544) and up (0.7005) movements, with comparatively higher neutral sensitivity (0.4581) than observed for Tesla or Apple, suggesting slightly improved recognition of stable periods for this asset.

Overall, the results indicate a consistent tendency across companies where the classifier is more effective at detecting directional price movements than neutral regimes, reflecting the inherent difficulty of modeling low-signal or sideways market conditions. This class-specific imbalance is typical in financial classification tasks and highlights the importance of evaluating performance beyond aggregate accuracy.

5. Discussions

This section discusses technical indicator-based feature importance from the extreme gradient boosting models for the three companies under study and provide comparisons. It also discusses interpretation of individual predictions in terms of technical indicators.

5.1 Feature Importance and Indicator Relevance for Tesla

Figure 2 provides a variable importance plot for Tesla that includes all 36 technical indicators used in this study.

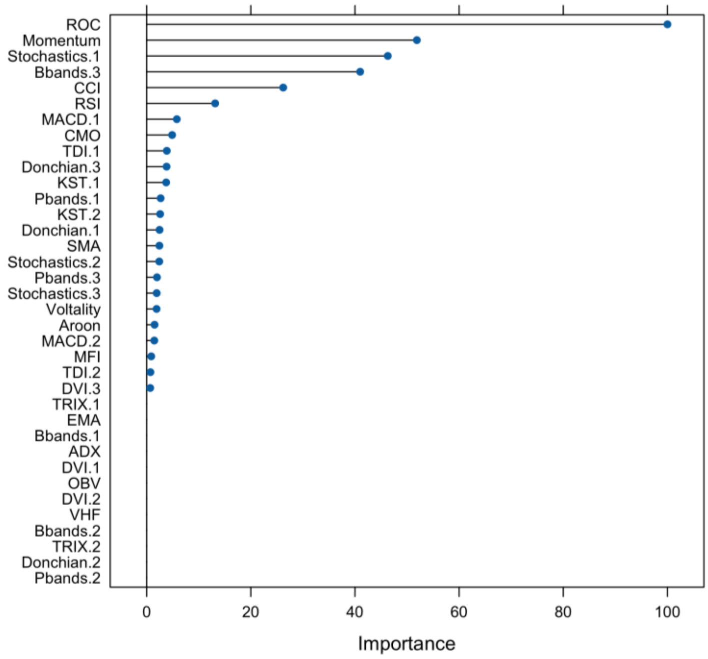

Fig. 2. Variable importance for Tesla

As seen from Figure 2, the variable importance profile for Tesla is dominated by Rate of Change (ROC), indicating that short-term price acceleration is the most influential signal for classifying directional movement. Momentum-based indicators—including Momentum, Stochastics, and Bollinger Bands (Bbands)—also exhibit substantial importance, highlighting Tesla's sensitivity to rapid trend continuation and overbought/oversold dynamics. Volatility-sensitive measures such as CCI and RSI contribute moderately, while longer-horizon trend indicators (e.g., EMA, ADX, OBV) show limited influence. Overall, the model emphasizes high-frequency momentum and volatility signals, consistent with Tesla's historically reactive and high-beta price behavior.

5.2 Feature Importance and Indicator Relevance for Apple

Figure 3 provides a variable importance plot for Apple that includes all 36 technical indicators used in this study.

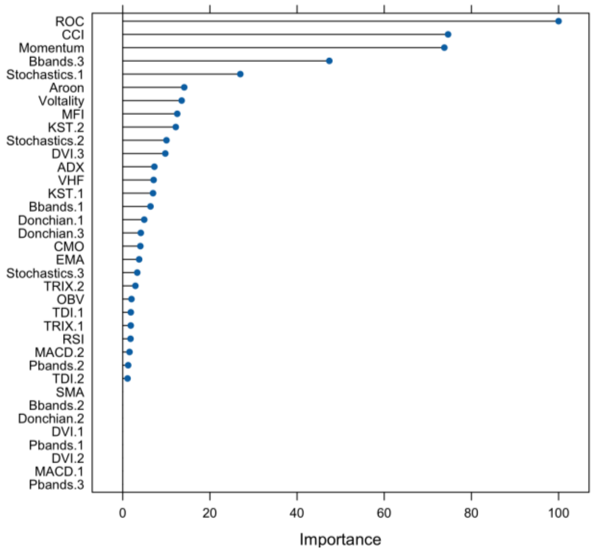

Fig. 3. Variable importance for Apple

As seen from Figure 3, Apple's importance distribution remains centered on ROC, but the ranking reveals a stronger role for trend-consistency indicators such as CCI, Momentum, and Bollinger Bands. Compared to Tesla, Apple exhibits a more balanced reliance on momentum and volatility, with secondary contributions from Aroon, MFI, and KST indicators. The reduced dominance of extreme momentum signals suggests that Apple's price dynamics are comparatively smoother, allowing the model to exploit structured trend confirmation rather than sharp directional bursts. Neutral and sideways regimes appear harder to distinguish, as reflected in the broader dispersion of moderately important indicators.

5.3 Feature Importance and Indicator Relevance for Nvidia

Figure 4 provides a variable importance plot for Nvidia that includes all 36 technical indicators used in this study.

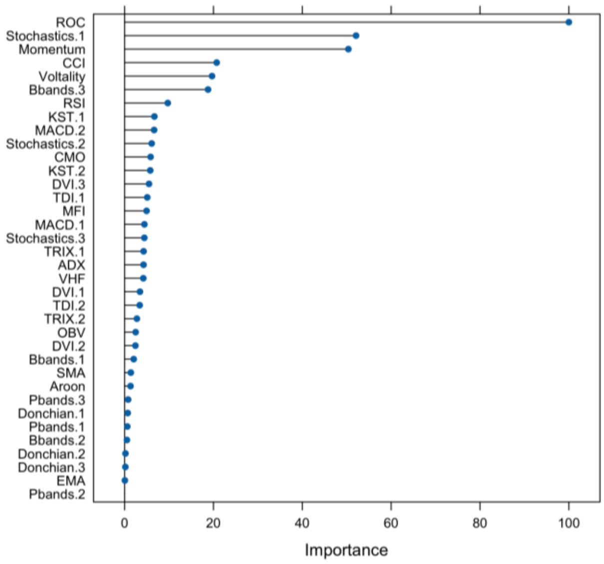

Fig. 4. Variable importance for Nvidia

As seen from Figure 4, for Nvidia, ROC again emerges as the single most influential predictor, but the subsequent importance pattern differs notably. Stochastic oscillators and Momentum indicators play a prominent role, indicating strong responsiveness to short-term cyclical movements. Volatility-related features (e.g., CCI, Volatility, Bollinger Bands) contribute meaningfully, while trend-strength measures such as KST and MACD provide incremental value. Compared to Tesla and Apple, Nvidia shows a more diversified importance distribution across momentum–volatility hybrids, reflecting its tendency toward sustained directional moves punctuated by sharp regime shifts.

Across all three assets, momentum-driven indicators—particularly ROC—consistently dominate, underscoring the central role of recent price changes in short-horizon stock movement classification. However, asset-specific importance profiles diverge beyond the top predictors. Tesla's model prioritizes aggressive momentum and volatility signals, Apple's model favors trend confirmation and stability, and Nvidia's model balances cyclical momentum with volatility-sensitive indicators. These differences suggest that while the directional signal is universal, the mechanisms through which it manifests are asset dependent, reinforcing the need for asset-specific modeling even under a common feature set.

5.4 Interpreting individual predictions and implications for Tesla

Figure 5 provides explanation of individual predictions for the final extreme gradient boosting model from the test data for Tesla.

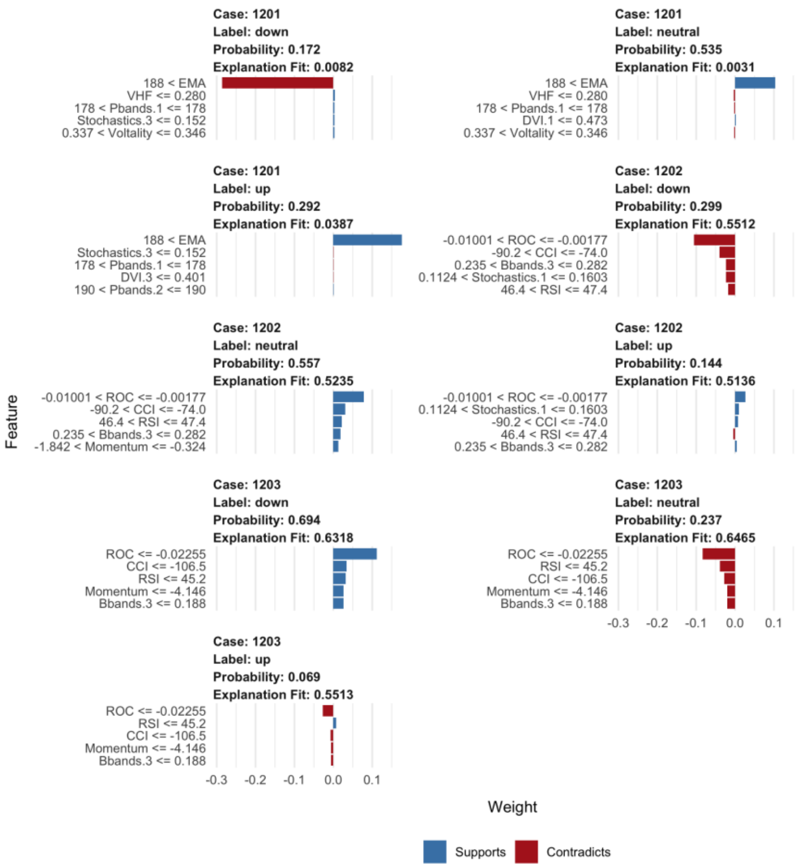

Fig. 5. Explaining individual predictions for Tesla

The Figure 5 presents case-level explanations for three representative Tesla observations from the test data (Cases 1201–1203), showing how individual technical indicators contribute to the predicted probabilities of down, neutral, and up price movements. Blue bars indicate features that support the predicted class, while red bars indicate features that contradict it. The explanation fit quantifies how well the linear surrogate approximates the local model behavior.

Note that Case 1201 predominantly belongs to neutral category. For Case 1201, the model assigns the highest probability to the neutral class (0.535), with substantially lower probabilities for up (0.292) and down (0.172). The neutral prediction is primarily supported by conditions involving EMA, VHF, price bands (Pbands), and DVI, suggesting a market state characterized by low directional strength and consolidation. Momentum and volatility signals are present but insufficiently strong to push the prediction toward a directional regime, indicating sideways price behavior.

In Case 1202, the model again favors the neutral class (0.557), although the explanation reveals stronger competing signals. Moderate ROC, CCI, RSI, and Stochastics values jointly support neutrality, reflecting mixed momentum and oscillatory behavior. The relatively high explanation fit indicates that these indicators collectively provide a coherent rationale for the neutral classification, despite weaker opposing signals toward the up or down classes.

Case 1203 exhibits a dominant down prediction (0.694), with very low probability assigned to the up class (0.069). The explanation shows strong support from negative ROC, deeply negative CCI, low RSI, and strongly negative Momentum, all of which are classical signatures of bearish momentum and oversold conditions. These indicators consistently reinforce the downward classification, yielding a high explanation fit and indicating a robust, unambiguous bearish signal.

Across all cases, the explanations demonstrate that Tesla's predictions are primarily driven by short-term momentum and oscillator-based indicators, with ROC, CCI, RSI, Stochastics, and volatility-adjusted measures playing central roles. Neutral classifications emerge when these signals are weak or conflicting, whereas strong alignment among momentum and oversold indicators produces confident directional predictions. This behavior aligns with Tesla's historically high-volatility, momentum-sensitive price dynamics, and provides transparent, case-specific justification for the model's decisions.

5.5 Interpreting individual predictions and implications for Apple

Figure 6 provides explanation of individual predictions for the final extreme gradient boosting model from the test data for Apple.

The Figure 6 reports case-level explanations for three representative Apple observations (Cases 1222–1224). For each case, the model's predicted probabilities for up, neutral, and down are shown alongside feature contributions. Blue bars indicate indicators that support a given class, while red bars indicate indicators that contradict it. The explanation fit reflects how well the local surrogate captures the model's behavior about the instance.

In Case 1222, the neutral class receives the highest probability (0.434), with down (0.384) and up (0.182) close behind, indicating uncertainty and weak directionality. Contributions from Momentum, Stochastics, Bollinger Bands, ROC, and OBV are small and partially offsetting, suggesting muted price acceleration and limited volume confirmation. The model therefore favors a sideways regime, consistent with mixed momentum and volatility cues.

For Case 1223, probabilities are split between neutral (0.487) and up (0.337), while down is less likely (0.176). Positive support for the up class comes from higher Stochastics, moderate positive Momentum, and positive ROC, indicating emerging upward pressure. However, simultaneous contradictory signals—particularly from Bollinger Bands and OBV—temper confidence, leading the

Applied Decision Analytics
Volume 2, Issue 1 (2026) 124-140

model to retain neutral as the most probable outcome. This pattern reflects incipient but unconfirmed trend formation.

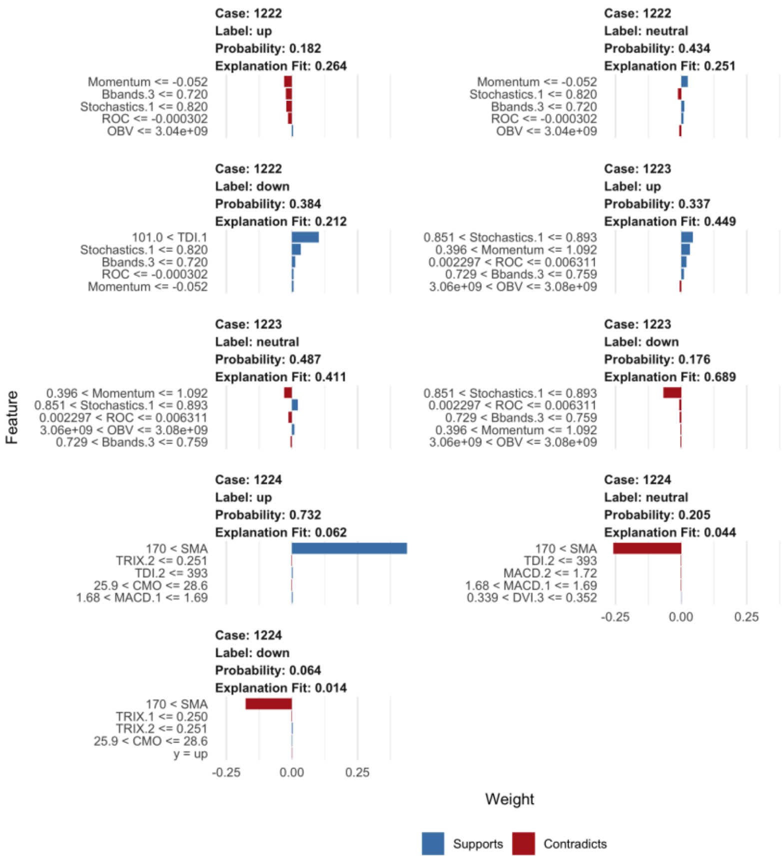

Fig. 6. Explaining individual predictions for Apple

Case 1224 shows a strong up prediction (0.732), with low probabilities for neutral (0.205) and down (0.064). The explanation highlights substantial positive contributions from SMA, TRIX, CMO, and MACD, all pointing to trend continuation and positive momentum. Contradictory evidence for the neutral or down classes is minimal, yielding a decisive bullish classification despite a relatively low explanation fit, which suggests dominance by a small number of strong indicators.

Across the Apple cases, the model relies on a combination of momentum and trend-confirmation indicators—notably Momentum, Stochastics, ROC, SMA, MACD, and TRIX—to differentiate regimes. Neutral predictions arise when these signals are weak or conflicting, whereas upward predictions emerge when multiple momentum and trend measures align. Compared with more volatile assets, Apple's explanations reflect smoother transitions between regimes and a greater emphasis on trend confirmation over abrupt momentum shifts, consistent with Apple's comparatively stable price dynamics.

5.6 Interpreting individual predictions and implications for Nvidia

Figure 7 provides explanation of individual predictions for the final extreme gradient boosting model from the test data for Nvidia.

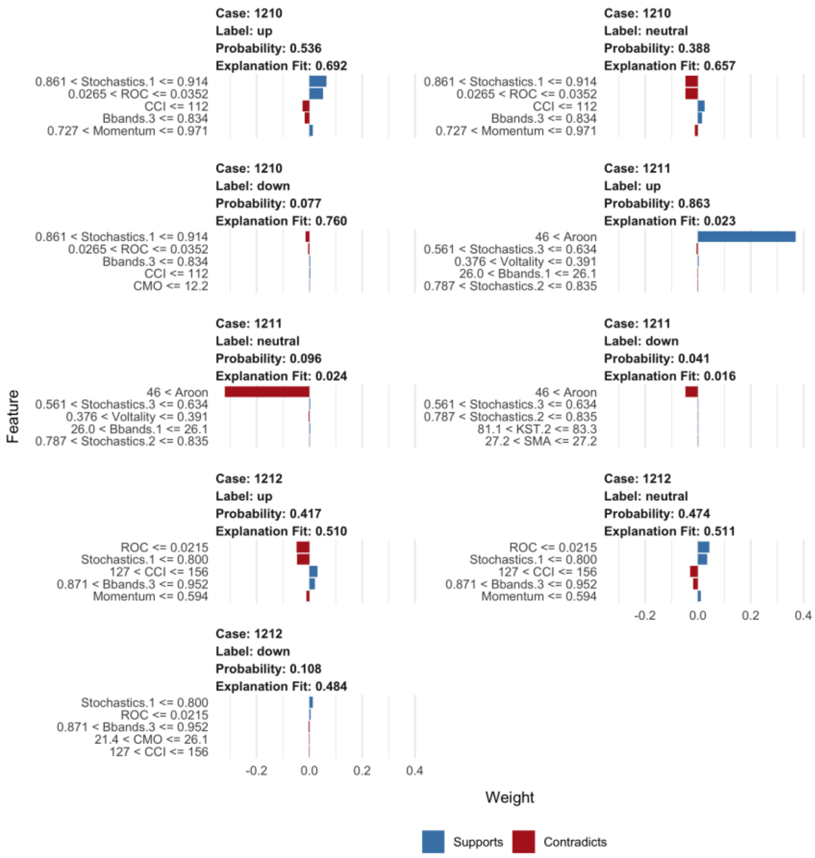

Fig. 7. Explaining individual predictions for Nvidia

The Figure 7 illustrates case-level explanations for three representative Nvidia observations (Cases 1210–1212). For each case, predicted probabilities for up, neutral, and down movements are shown alongside feature contributions. Blue bars denote indicators that support a given class, while red bars denote indicators that contradict it. The explanation fit measures how faithfully the local surrogate approximates the underlying XGBoost decision behavior.

For Case 1210, the model assigns the highest probability to the up class (0.536), followed by neutral (0.388), with minimal support for down (0.077). The upward prediction is primarily supported by high Stochastics, positive ROC, and strong Momentum, indicating short-term acceleration and bullish oscillatory behavior. However, moderate counter-signals from CCI and Bollinger Bands temper confidence, leaving non-trivial probability mass on the neutral class. This reflects a bullish but not fully confirmed trend.

Case 1211 exhibits a highly confident up prediction (0.863), with negligible probabilities for neutral (0.096) and down (0.041). The explanation shows dominant support from Aroon, Stochastics, Volatility, and Bollinger Bands, collectively signaling trend emergence, directional strength, and expanding price range. The very low explanation fit for competing classes indicates that contradictory evidence is weak, yielding a clear and decisive bullish classification.

In Case 1212, probabilities are split between neutral (0.474) and up (0.417), with down remaining unlikely (0.108). Feature contributions reveal mixed signals: moderate ROC and Stochastics provide limited upward support, while CCI, CMO, and Bollinger Bands introduce conflicting evidence. The similar explanation fit for up and neutral classes indicates regime uncertainty, consistent with a transition phase where directional momentum is present but insufficiently strong to dominate.

Across the Nvidia cases, predictions are driven primarily by momentum- and trend-sensitivity indicators, notably Stochastics, ROC, Aroon, Momentum, and volatility-adjusted measures. Strong alignment among these indicators produces high-confidence upward predictions, while mixed or moderate signals lead to neutral classifications. Compared with Tesla and Apple, Nvidia's explanations show a pronounced sensitivity to trend-emergence indicators (Aroon, Stochastics), reflecting its tendency toward sustained directional moves once momentum establishes.

5.7 Linking Local Explanations to Class-Wise Sensitivity

The class-wise sensitivity results align closely with the local explanation patterns observed across assets, providing a coherent interpretation of model behavior. For Tesla, high sensitivity for down and up classes, combined with lower sensitivity for neutral, is consistent with local explanations showing strong alignment among momentum (ROC, Momentum) and oscillator-based indicators (CCI, RSI, Stochastics) during directional regimes. When these indicators reinforce one another, the model produces confident directional predictions, yielding higher recall for down and up movements. Conversely, neutral cases exhibit weaker or conflicting signals, explaining Tesla's comparatively lower neutral sensitivity.

For Apple, class-wise sensitivity is highest for down, moderate for up, and lowest for neutral, which mirrors the explanation results. Local explanations reveal that Apple's neutral predictions often arise from partially offsetting momentum and trend-confirmation signals (e.g., mixed Momentum, ROC, and Bollinger Band cues), making neutral regimes harder to detect reliably. In contrast, downward movements are supported by clearer combinations of negative momentum and oversold indicators, resulting in stronger recall for the down class. Upward predictions depend more heavily on trend-confirmation indicators (SMA, MACD, TRIX), which emerge less frequently, explaining the slightly lower sensitivity for up relative to down.

For Nvidia, relatively balanced and comparatively higher sensitivity across down, neutral, and up classes is consistent with its explanation structure. Local explanations show that Nvidia's predictions rely on a diversified set of momentum–trend indicators, including Stochastics, ROC, Aroon, volatility measures, and Bollinger Bands, which provide clearer separation even during transitional regimes. This richer and more consistent signal structure improves the model's ability to correctly identify neutral states, explaining Nvidia's superior neutral sensitivity relative to Tesla and Apple.

Taken together, the sensitivity metrics and local explanations indicate that directional classes (down/up) achieve higher recall when momentum and oscillator signals align strongly, while neutral regimes suffer lower sensitivity when signals are weak or conflicting. Asset-specific differences in price dynamics amplify this effect: Tesla's high volatility favors strong directional recall, Apple's smoother trends obscure neutrality, and Nvidia's pronounced trend-emergence behavior improves class balance. These findings validate that the observed class-wise sensitivities are a direct consequence of the indicator-level decision logic revealed by local explanations, rather than artifacts of model fitting.

4. Conclusions

This study examined the effectiveness and interpretability of XGBoost-based multi-class stock movement prediction models built exclusively on technical indicators for Tesla, Apple, and Nvidia. By combining rigorous two-stage hyperparameter tuning, out-of-sample evaluation, and both global and local explainability analyses, the study addressed key gaps in prior research on financial machine learning.

Consistent with the first research objective, the results show that XGBoost models achieve strong out-of-sample performance for directional price movements, with higher sensitivity for up and down classes than for neutral regimes across all three equities. This confirms that pronounced market movements are more predictable than low-signal consolidation periods. Nvidia exhibited the most balanced class-wise performance, while Tesla and Apple showed stronger asymmetries favoring directional predictions.

Addressing the second objective, feature importance analysis revealed that momentum-based indicators, particularly Rate of Change (ROC), are the dominant predictors across assets, while secondary indicators differ by firm. Tesla relies more heavily on aggressive momentum and volatility signals, Apple on trend-consistency measures, and Nvidia on a diversified mix of momentum and trend-emergence indicators. These findings highlight the importance of asset-specific indicator relevance, even under a common modeling framework.

In line with the third objective, local explanation analysis demonstrated that class-wise sensitivity differences are directly explained by indicator-level decision logic. Directional predictions arise when momentum and oscillator signals align strongly, whereas neutral classifications occur under weak or conflicting signals. This linkage confirms that observed performance asymmetries reflect meaningful market structure rather than opaque model behavior.

Several limitations warrant attention. The analysis focuses on three large-cap U.S. equities, excludes fundamental and macroeconomic variables, and relies on fixed thresholds for class definition. Moreover, economic considerations such as transaction costs are not incorporated. Future research could extend this framework to broader asset universes, integrate additional data sources, explore adaptive class definitions, and evaluate trading performance under realistic market constraints.

Overall, the study demonstrates that interpretable, well-tuned ensemble models can extract economically meaningful signals from technical indicators, while also clarifying the conditions under which prediction remains inherently difficult.

Acknowledgement
This research was not funded by any grant.

Conflicts of Interest
The authors declare no conflicts of interest.

References

[1] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. The Review of Financial Studies, 33(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

[2] Shi, C. (2025). From econometrics to machine learning: Transforming empirical asset pricing. Journal of Economic Surveys. Advance online publication. https://doi.org/10.1111/joes.70002

[3] Ye, J., Goswami, B., Gu, J., Uddin, A., & Wang, G. (2024). From factor models to deep learning: Machine learning in reshaping empirical asset pricing [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2403.06779

[4] Rai, B., & Soltanisehat, L. (2025). A Multi-Model Machine Learning Framework for Daily Stock Price Prediction. *Big Data and Cognitive Computing*, *9*(10), 248. https://doi.org/10.3390/bdcc9100248

[5] Cohen, G., Aiche, A., & Eichel, R. (2025). Artificial intelligence models for predicting stock returns using fundamental, technical, and entropy-based strategies: A semantic-augmented hybrid approach. Entropy, 27(6), 550. https://doi.org/10.3390/e27060550

[6] Fozap, F. M. P. (2025). Hybrid machine learning models for long-term stock market forecasting: Integrating technical indicators. Journal of Risk and Financial Management, 18(4), Article 201. https://doi.org/10.3390/jrfm18040201

[7] Saud, A. S., & Shakya, S. (2024). Technical indicator empowered intelligent strategies to predict stock trading signals. Journal of Open Innovation: Technology, Market, and Complexity, 10(4), Article 100398. https://doi.org/10.1016/j.joitmc.2024.100398

[8] Rezaei, H., Hooshmand, M., & Abdi, K. (2025). Machine learning based classification and regression approaches for return and risk prediction in the stock market using fundamental feature engineering. Discover Analytics, 3, Article 17. https://doi.org/10.1007/s44257-025-00042-5

[9] Balasubramanian, P., C., P., Badarudeen, S., & Sriraman, H. (2024). A systematic literature survey on recent trends in stock market prediction. PeerJ Computer Science, 10, Article e1700. https://doi.org/10.7717/peerj-cs.1700

[10] Khan, A. H., Shah, A., Ali, A., Shahid, R., Zahid, Z. U., Sharif, M. U., Jan, T., & Zafar, M. H. (2023). A performance comparison of machine learning models for stock market prediction with novel investment strategy. PLoS ONE, 18(9), Article e0286362. https://doi.org/10.1371/journal.pone.0286362

[11] Uddin, S., & Lu, H. (2024). Confirming the statistically significant superiority of tree-based machine learning algorithms over their counterparts for tabular data. PLoS ONE, 19(4), Article e0301541. https://doi.org/10.1371/journal.pone.0301541

[12] Patsiarikas, M., Papageorgiou, G., & Tjortjis, C. (2025). Using machine learning on macroeconomic, technical, and sentiment indicators for stock market forecasting. Information, 16(7), Article 584. https://doi.org/10.3390/info16070584

[13] Liu, J. (2024). Predicting Chinese stock market using XGBoost multi-objective optimization with optimal weighting. PeerJ Computer Science, 10, e1931. https://doi.org/10.7717/peerj-cs.1931

[14] Indra, Supian, S., Sukono, Riaman, Saptura, M., Azahra, A., & Pirdaus, D. (2025). Stock return prediction on the LQ45 market index in the Indonesia Stock Exchange using a machine learning algorithm based on technical indicators. Journal of Risk and Financial Management, 18(12), Article 714. https://doi.org/10.3390/jrfm18120714

[15] Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on tabular data? NeurIPS 2022 Dataset and Benchmark Track. https://arxiv.org/abs/2207.08815

[16] Shmuel, A., Glickman, O., & Lazebnik, T. (2025). A comprehensive benchmark of machine and deep learning models on structured data for regression and classification. Neurocomputing, 655, Article 131337. https://doi.org/10.1016/j.neucom.2025.131337

[17] Zhang, T., Huo, M. D., Ma, Z., Hu, J., Liang, Q., & Chen, H. (2024). Prediction model of stock return on investment based on hybrid DNN and TabNet model. PeerJ Computer Science, 10, Article e2057. https://doi.org/10.7717/peerj-cs.2057

[18] Chicco, D., & Jurman, G. (2022). The ABC recommendations for validation of supervised machine learning results in biomedical sciences. Frontiers in Big Data, 5, Article 979465. https://doi.org/10.3389/fdata.2022.979465

[19] Olorunnimbe, K., & Viktor, H. L. (2023). Deep learning in the stock market—a systematic survey of practice, backtesting, and applications. Artificial Intelligence Review, 56(Suppl 3), 3159–3210. https://doi.org/10.1007/s10462-022-10226-0

[20] Prata, M., Masi, G., Berti, L. et al. (2024). LOB-based deep learning models for stock price trend prediction: A benchmark study. Artificial Intelligence Review, 57, Article 116. https://doi.org/10.1007/s10462-024-10715-4

[21] Vancsura, L., Tatay, T., & Bareith, T. (2025). Navigating AI-Driven Financial Forecasting: A Systematic Review of Current Status and Critical Research Gaps. *Forecasting*, *7*(3), 36. https://doi.org/10.3390/forecast7030036

[22] Karulkar, Y., Shah, A., & Naik, R. (2025). From data to decisions: Evaluating machine learning models for stock market forecasting. Vision: The Journal of Business Perspective. Advance online publication. https://doi.org/10.1177/09711023251349445