African Scientific Journal
ISSN : 2658-9311
Vol : 03, Num 35 Avril 2026

Machine Learning in Finance: An Explainable Multi-Model Framework for Stock Return Forecasting and Risk-Aware Portfolio Allocation.

Auteur 1 : Youssef LAHBABI.
Auteur 2 : Ouissam EL ALAOUI.
Auteur 3 : Farid OUBAALI.

LAHBABI Youssef
EMAA Business School – Agadir –Morocco

OUISSAM EL ALAOUI
EMAA Business School – Agadir –Morocco

OUBAALI Farid
EMAA Business School – Agadir –Morocco

Déclaration de divulgation : L'auteur n'a pas connaissance de quelconque financement qui pourrait affecter l'objectivité de cette étude.

Conflit d'intérêts : L'auteur ne signale aucun conflit d'intérêts.

Pour citer cet article : LAHBABI .Y, EL ALAOUI .O & OUBAALI .F (2026) « Machine Learning in Finance: An Explainable Multi-Model Framework for Stock Return Forecasting and Risk-Aware Portfolio Allocation », African Scientific Journal « Volume 03, Num 35 » pp: 2976– 3000.

DOI : 10.5281/zenodo.20230041

African Scientific Journal
ISSN : 2658-9311
Vol : 03, Num 35 Avril 2026

Abstract

Machine learning is an essential methodology in modern quantitative finance, given its capacity to capture nonlinear relationships, temporal dependencies, and high-dimensional predictor structures that are not easily modeled using traditional econometric techniques. This paper describes an explainable multi-model framework for stock return forecasting and risk-aware portfolio allocation. The framework combines Ordinary Least Squares (OLS), a linear benchmark model, with advanced machine learning models such as Random Forest, Extreme Gradient Boosting (XGBoost), and Long Short-Term Memory (LSTM) networks. Such a combination allows thorough comparisons between linear, nonlinear, and sequential modeling approaches. An empirical analysis is conducted using real-world financial and macro-financial data on daily equity prices, trading activity indicators, the CBOE Volatility Index (VIX), and Fama–French factors, to validate the proposed approach. To enhance interpretability and transparency of results, SHAP (SHapley Additive exPlanations) has been adopted in this framework for global as well as local interpretation of feature contributions toward model outputs. This enables deeper economic insights into predictive mechanisms rather than relying solely on statistical performance. Apart from the standard forecast evaluation, this study establishes a direct link between the predictive outputs and the economic decisions by inserting the forecasts into a dynamic portfolio allocation strategy. A detailed validation framework is introduced here that includes not only predictive accuracy but also directional (sign) accuracy along with risk-adjusted economic performance metrics.

The main contribution of this research is bringing together predictive performance, model explainability, and portfolio-level economic relevance within one empirical framework. It advances the literature by connecting machine learning-based forecasting with its real-world application in asset management to enable AI-assisted financial analytics and decision-making systems.

This study adopts a comparative empirical design combining Ordinary Least Squares (OLS), Random Forest, XGBoost, and Long Short-Term Memory (LSTM) models to forecast stock returns and support risk-aware portfolio allocation. Sample and data: The empirical analysis relies on real-world daily financial and macro-financial data, including adjusted stock prices, trading volume, the CBOE Volatility Index (VIX), and Fama–French factors. Main findings: The proposed explainable multi-model framework improves the evaluation of stock return forecasting by combining predictive accuracy, directional accuracy, SHAP-based interpretability, and portfolio-level risk-adjusted performance. The results suggest that machine learning models, particularly XGBoost and LSTM, can provide economically meaningful signals for dynamic portfolio allocation when combined with appropriate risk constraints.

Keywords: Machine Learning; Finance; Explainable AI; Stock Return Forecasting; XGBoost; LSTM; Portfolio Allocation; Risk Management.

List of Abbreviations:

AI = Artificial Intelligence

OLS = Ordinary Least Squares

LSTM = Long Short-Term Memory

XGBoost = Extreme Gradient Boosting

SHAP = SHapley Additive exPlanations

VIX = Volatility Index

RMSE = Root Mean Squared Error

MAE = Mean Absolute Error

Résumé

Le machine learning constitue aujourd'hui une méthodologie essentielle en finance quantitative moderne, en raison de sa capacité à capturer des relations non linéaires, des dépendances temporelles ainsi que des structures prédictives de grande dimension, difficilement modélisables à l'aide des techniques économétriques traditionnelles. Cet article propose un cadre explicable multi-modèles destiné à la prévision des rendements boursiers et à l'allocation de portefeuille sensible au risque. Le cadre combine les Moindres Carrés Ordinaires (OLS) utilisés comme modèle linéaire de référence — avec des modèles avancés de machine learning tels que Random Forest, Extreme Gradient Boosting (XGBoost) et les réseaux de neurones Long Short-Term Memory (LSTM). Cette combinaison permet d'effectuer des comparaisons approfondies entre des approches de modélisation linéaires, non linéaires et séquentielles.

Une analyse empirique est menée à partir de données financières et macro-financières réelles incluant les prix quotidiens des actions, les indicateurs d'activité de trading, l'indice de volatilité CBOE (VIX) ainsi que les facteurs de Fama–French, afin de valider l'approche proposée. Afin d'améliorer l'interprétabilité et la transparence des résultats, la méthode SHAP (SHapley Additive exPlanations) est intégrée au cadre méthodologique pour fournir des interprétations globales et locales des contributions des variables aux prédictions des modèles. Cela permet d'obtenir des analyses économiques plus approfondies des mécanismes prédictifs, au-delà des seules performances statistiques.

En plus de l'évaluation standard des prévisions, cette étude établit un lien direct entre les résultats prédictifs et les décisions économiques en intégrant les prévisions dans une stratégie dynamique d'allocation de portefeuille. Un cadre détaillé de validation est introduit, incluant non seulement la précision prédictive, mais également la précision directionnelle (du signe) ainsi que des indicateurs de performance économique ajustés au risque. La principale contribution de cette recherche réside dans l'intégration, au sein d'un même cadre empirique, de la performance prédictive, de l'explicabilité des modèles et de la pertinence économique au niveau du portefeuille. Cette étude fait progresser la littérature en reliant les techniques de prévision basées sur le machine learning à leurs applications réelles dans la gestion d'actifs, contribuant ainsi au développement de systèmes analytiques et décisionnels financiers assistés par l'intelligence artificielle.

Cette étude adopte un dispositif empirique comparatif combinant les Moindres Carrés Ordinaires (OLS), Random Forest, XGBoost et les réseaux Long Short-Term Memory (LSTM), afin de prévoir les rendements boursiers et d'appuyer une allocation de portefeuille sensible au risque. Échantillon et données : L'analyse empirique repose sur des données financières et macro-financières quotidiennes réelles, incluant les prix ajustés des actions, le volume de transaction, l'indice de volatilité CBOE (VIX) ainsi que les facteurs de Fama–French. Principaux résultats : Le cadre multi-modèles explicable proposé améliore l'évaluation de la prévision des rendements boursiers en combinant précision prédictive, précision directionnelle, interprétabilité fondée sur SHAP et performance économique ajustée au risque au niveau du portefeuille. Les résultats suggèrent que les modèles de machine Learning, notamment XGBoost et LSTM, peuvent fournir des signaux économiquement significatifs pour une allocation dynamique de portefeuille lorsqu'ils sont associés à des contraintes de risque appropriées.

Mots-clés : Machine Learning ; Finance ; Intelligence Artificielle Explicable ; Prévision des Rendements Boursiers ; XGBoost ; LSTM ; Allocation de Portefeuille ; Gestion des Risques.

Introduction

Financial markets are characterized by nonlinear dynamics, volatility clustering, abrupt regime shifts, and high sensitivity to macroeconomic uncertainty. These stylized facts reduce the effectiveness of purely linear approaches when the objective is to forecast expected returns and support investment decisions under uncertainty. In this context, machine learning has emerged as a powerful methodological alternative because of its ability to model nonlinear interactions, high-dimensional structures, and time-varying relationships that conventional econometric approaches often fail to capture adequately (Gu et al., 2020; Fischer & Krauss, 2018).

The relevance of machine learning in finance has increased substantially in recent years, particularly in empirical asset pricing, market forecasting, and risk management. Gu et al. (2020) show that machine learning methods can generate significant gains in empirical asset pricing relative to standard linear models, mainly because these techniques are able to identify complex and nonlinear predictor interactions. Similarly, Fischer and Krauss (2018) demonstrate that Long Short-Term Memory (LSTM) networks can outperform several traditional prediction approaches in financial market forecasting, especially in environments characterized by temporal dependence and noisy signals. These findings confirm that the integration of machine learning into finance is no longer marginal but central to modern quantitative research.

At the same time, better predictive performance does not automatically imply better financial decision-making. A forecasting model may reduce statistical error while remaining unstable across regimes, opaque in its decision logic, or economically weak once translated into portfolio allocation. This methodological tension has become increasingly important in recent research, where explainability and practical usability are now viewed as key dimensions of trustworthy financial artificial intelligence. In that regard, SHAP-based interpretation has gained growing importance because it helps explain how each predictor contributes to model output, thereby improving transparency in black-box forecasting systems (Lundberg & Lee, 2017).

To strengthen the empirical foundation of machine learning models in finance, this study relies on real and recognized public financial sources. Market uncertainty is proxied by the CBOE Volatility Index (VIX), which FRED describes as a daily indicator of market expectations of near-term volatility conveyed by stock index option prices. In parallel, the study incorporates factor-based variables from the Kenneth R. French Data Library, where the Fama–French factors are officially described as returns constructed from value-weight portfolios formed on size and book-to-market characteristics. These sources provide a robust empirical basis for integrating both market dynamics and systematic risk signals into the predictive framework (Federal Reserve Bank of St. Louis, n.d.; French, n.d.).

Accordingly, this article addresses the following research question: Can an explainable multi-model machine learning framework improve stock return forecasting while also generating economically meaningful and risk-aware portfolio allocation decisions? To answer this question, the study compares a linear benchmark with Random Forest, XGBoost, and LSTM models using real market and macro-financial data. The originality of the article lies in combining three dimensions that are still too rarely integrated in a single framework: predictive performance, explainability, and portfolio-level economic relevance. Such integration is especially important because the African Scientific Journal requires theoretical rigor, a clear research object, and an up-to-date bibliography under APA standards.

The study makes four main contributions. First, it provides a unified comparative framework across linear, tree-based, and sequence-based models. Second, it integrates explainability through SHAP-based feature attribution. Third, it evaluates the usefulness of predictions not only through statistical error metrics but also through a dynamic portfolio allocation layer under risk constraints. Fourth, it proposes a publication-oriented structure aligned with the editorial expectations of the journal and with recent developments in finance and machine learning research.

1. Literature Review and Research Gap

1.1. Machine Learning and Financial Forecasting

The rapid evolution of financial markets, combined with the increasing availability of high-frequency and large-scale financial data, has accelerated the integration of machine learning techniques into quantitative finance. Traditional econometric models, such as linear regression and autoregressive frameworks, remain useful in financial analysis; however, they often struggle to capture nonlinear relationships, complex variable interactions, and regime-dependent market behavior. In this context, machine learning provides a more flexible methodological framework for financial forecasting because it can model nonlinear patterns and extract predictive signals from large and heterogeneous datasets (Gu et al., 2020).

Recent empirical research confirms that machine learning has become particularly influential in stock return prediction, volatility forecasting, asset pricing, credit risk assessment, and portfolio management. Gu et al. (2020) demonstrate that machine learning techniques can improve empirical asset pricing by capturing nonlinear relationships between predictors and expected returns. Their work shows that flexible models such as tree-based algorithms and neural networks may generate stronger out-of-sample performance than traditional linear approaches when applied to financial prediction problems.

Among machine learning models, tree-based methods such as Random Forest and gradient boosting have received considerable attention because of their robustness, ability to manage noisy predictors, and capacity to model complex interactions between financial variables. XGBoost, introduced by Chen and Guestrin (2016), is especially relevant in financial forecasting because it combines predictive accuracy, scalability, and regularization mechanisms that reduce overfitting. This makes it suitable for structured financial datasets where predictors may include lagged returns, volatility indicators, technical variables, macro-financial signals, and factor-based variables.

Deep learning methods have also gained importance in financial forecasting, especially when the objective is to model sequential dependencies in time-series data. Long Short-Term Memory networks are widely used because they can learn temporal patterns and delayed effects that are difficult to capture using static models. Fischer and Krauss (2018) show that LSTM networks can outperform several conventional approaches in financial market prediction, particularly when historical price information contains temporal structures useful for forecasting future movements.

However, despite these advances, financial forecasting remains a difficult task because market signals are noisy, unstable, and often weak. Predictive models may perform well during one market regime but lose accuracy during periods of stress, high volatility, or structural change. This limitation highlights the importance of using robust validation procedures, such as chronological train-test splits, rolling-window estimation, and out-of-sample performance evaluation. In this sense, the quality of a financial machine learning model should not be assessed only through in-sample accuracy, but also through its ability to generalize across different market conditions (Fischer & Krauss, 2018; Gu et al., 2020).

Another important limitation in the literature is that many machine learning studies focus mainly on statistical accuracy while giving less attention to economic usefulness. A model may reduce forecast errors without improving portfolio performance or risk-adjusted returns. Therefore, recent finance-oriented machine learning research increasingly emphasizes the need to connect prediction with investment decision-making. This means evaluating models not only through RMSE, MAE, or directional accuracy, but also through financial indicators such as Sharpe ratio, maximum drawdown, turnover, and risk-adjusted cumulative returns (Gu et al., 2020).

From this perspective, the present study contributes to the literature by proposing a comparative framework that combines linear models, tree-based machine learning models, and sequence-based deep learning models for stock return forecasting. Unlike studies that focus only on predictive comparison, this article links forecasting outputs to a risk-aware portfolio allocation layer. This design allows the research to assess whether machine learning forecasts are not only statistically accurate but also economically meaningful.

The literature on machine learning in finance has expanded rapidly because financial returns are difficult to model using linear assumptions alone. Financial signals are often noisy, weak, unstable over time, and highly dependent on interactions across market variables. Recent review work synthesizing 187 Scopus-indexed studies from 2020 to 2024 shows that deep learning has become a major paradigm in financial forecasting, especially for return prediction, volatility estimation, and time-series modeling. However, that same review notes that a large part of the literature remains descriptive, fragmented, or weakly connected to actionable decision systems.

1.2. Explainable Artificial Intelligence in Financial Forecasting

As machine learning becomes more powerful, model opacity becomes more problematic. In finance, this issue is especially important because high-stakes decisions often require transparency, governance, and auditability. The rapid adoption of explainable artificial intelligence (XAI) reflects this need. Explainability has become increasingly important in financial forecasting, algorithmic trading, credit scoring, portfolio management, and regulatory analytics because financial institutions must justify and interpret automated decisions (Yeo et al., 2025).

Among explainability techniques, SHAP (SHapley Additive exPlanations) has become one of the most widely used approaches because it provides additive feature attribution that can be interpreted both locally and globally. Lundberg and Lee (2017) demonstrate that SHAP offers a unified framework for interpreting machine learning predictions by assigning contribution values to each predictor variable. This makes SHAP particularly attractive for financial forecasting systems where users need to understand why a predictive signal was generated and which variables contributed most to the decision process.

The increasing importance of explainable AI in finance is also reflected in recent surveys and scientific discussions that frame interpretability as a necessary dimension of trustworthy financial artificial intelligence. Recent literature emphasizes that explainability is no longer considered optional in financial AI systems because opaque models may create governance challenges, regulatory concerns, and difficulties in model validation (Arsenault et al., 2025; Yeo et al., 2025). Consequently, explainable machine learning is becoming an essential component of modern quantitative finance, especially in environments where predictive accuracy must be balanced with transparency and accountability.

1.3. Real Data Foundations for Financial ML

A robust finance–ML article requires credible data infrastructure. In this study, the macro-financial uncertainty proxy is the CBOE Volatility Index available through FRED. FRED identifies the VIXCLS series as the CBOE Volatility Index, reported at a daily frequency and used as a market-based volatility indicator (Federal Reserve Bank of St. Louis, n.d.). The factor side of the design relies on the Kenneth R. French Data Library, whose official documentation describes the Fama/French factors as constructed from value-weight portfolios formed on size and book-to-market characteristics (French, n.d.). The explanatory variables used in this study, namely lagged returns, volatility measures, momentum indicators, trading volume, VIX, and Fama–French factors, are selected based on prior empirical finance and machine learning studies, particularly Fama and French (1993), Fischer and Krauss (2018), and Gu et al. (2020).

1.4. Research Gap

Despite these advances, a significant gap remains. Many studies compare machine learning models using forecasting error metrics alone, without asking whether forecast improvements translate into superior investment outcomes once risk is taken into account. Gu et al. (2020) emphasize that the value of machine learning in finance should also be assessed through economic gains and investor-relevant performance, not only through statistical prediction accuracy. Likewise, many studies adopt complex models without incorporating an explanation layer that would make those models interpretable in practice. This limitation is important because SHAP-based explainability provides a unified framework for interpreting model predictions and assigning feature-level contributions to complex machine learning outputs (Lundberg & Lee, 2017). The underexplored intersection therefore lies in the simultaneous integration of prediction, explainability, and economic utility. This article addresses precisely that missing link.

2. Data and Methodology

2.1. Data Sources and Sample Description

This study relies on real financial and macro-financial data in order to ensure empirical credibility and reproducibility. The first data layer consists of daily stock market information, including adjusted closing prices, trading volume, and return-based indicators. The second layer includes the CBOE Volatility Index, obtained from FRED under the VIXCLS series. FRED reports VIXCLS as a daily CBOE Volatility Index series, which makes it suitable as a proxy for market uncertainty and volatility expectations (Federal Reserve Bank of St. Louis, n.d.). The third data layer relies on the Kenneth R. French Data Library. This source provides factor-based variables commonly used in empirical asset pricing, including market, size, and value-related factors. The official documentation indicates that the Fama/French factors are constructed from value-weight portfolios formed on size and book-to-market characteristics (French, n.d.). These data are widely used in quantitative finance and provide a strong factor-based foundation for the predictive framework (Fama & French, 1993; Gu et al., 2020).

2.2. Variable Construction

The dependent variable is the one-step-ahead daily stock return. It is calculated as the logarithmic change in adjusted closing prices:

$$R_{t+1} = \ln\left(\frac{P_{t+1}}{P_t}\right)$$

where P_t represents the adjusted closing price at time t , and R_{t+1} denotes the next-day return to be forecasted. The explanatory variables include lagged returns, rolling volatility, moving-average spreads, momentum indicators, volume variation, VIX, and Fama–French factors. These variables are selected because they capture different dimensions of financial market behavior: short-term price dynamics, risk conditions, trend effects, trading intensity, macro-financial uncertainty, and systematic risk exposure.

2.3. Feature Engineering

Feature engineering is a crucial step in financial machine learning because raw prices are rarely sufficient for robust forecasting. The study constructs several predictive variables from the original data. Lagged returns are used to capture short-term persistence or reversal. Rolling volatility is computed over a 20-day window to approximate recent risk conditions. Moving-average spreads are calculated as the difference between short-term and medium-term moving averages in order to capture trend dynamics. Momentum variables are added to represent cumulative price movements over recent periods. Volume variation is included as a proxy for market participation and liquidity.

The VIX variable is incorporated to account for market-wide uncertainty, while Fama–French factors are added to control for systematic sources of return variation. This combination allows the model to integrate both technical and factor-based information within a unified forecasting framework.

2.4. Machine Learning Models

The empirical analysis compares four predictive models. The first model is an Ordinary Least Squares regression, used as a transparent linear benchmark. The second model is Random Forest, which is suitable for nonlinear relationships and interaction effects. The third model is XGBoost, a gradient boosting method known for its predictive performance, scalability, and regularization capacity (Chen & Guestrin, 2016). The fourth model is Long Short-Term Memory, a deep learning architecture designed to model temporal dependencies in sequential data. LSTM models are particularly relevant in financial forecasting because they can capture delayed effects and memory structures in market time series (Fischer & Krauss, 2018).

The inclusion of these models allows the study to compare linear, tree-based, boosting-based, and sequence-based forecasting approaches within the same empirical design.

2.5. Model Evaluation

Model performance is assessed using both statistical and economic criteria. Statistical evaluation includes Root Mean Squared Error, Mean Absolute Error, out-of-sample R^2, and directional accuracy. These indicators measure the quality of the return forecasts from different perspectives.

However, because forecasting accuracy alone is not sufficient in finance, the study also evaluates economic usefulness. Predicted returns are transformed into dynamic portfolio weights using a risk-aware allocation rule. The resulting strategies are then assessed through Sharpe ratio, Sortino ratio, cumulative returns, maximum drawdown, and turnover-adjusted performance. This approach follows the view that machine learning models in finance should be judged not only by prediction errors but also by their ability to generate economically meaningful investment outcomes (Gu et al., 2020).

2.6. Explainability Procedure

To improve model transparency, this study incorporates an explainability layer based on SHAP values. SHAP provides a unified framework for interpreting model predictions by assigning each feature an importance value for a given prediction (Lundberg & Lee, 2017). This is especially relevant in finance, where users must understand whether predictions are driven by volatility, momentum, factor exposure, or market uncertainty.

Through this explainability procedure, the study evaluates not only which model performs best, but also which variables contribute most strongly to the predictive process. This makes the framework more suitable for financial decision-making, governance, and model validation.

3. Empirical Implementation and Validation Strategy

3.1. Data Preprocessing

Before model estimation, the collected datasets are cleaned and synchronized to ensure consistency across all variables. Missing observations are removed or aligned according to trading dates in order to avoid temporal inconsistencies between stock prices, volatility indicators, and factor-based variables. Financial time series frequently contain noise, extreme values, and irregularities; therefore, preprocessing is necessary to improve model stability and predictive reliability.

The adjusted closing price is selected instead of the raw closing price because it accounts for dividends and stock splits, making it more appropriate for return computation and long-horizon financial analysis. Daily returns are then calculated using logarithmic transformation because log returns are additive over time and commonly used in empirical finance research.

To improve the comparability of variables, several features are normalized or standardized before being introduced into machine learning models. This step is especially important for neural architectures such as LSTM, where large differences in variable scales may negatively affect convergence and learning quality.

3.2. Training and Testing Procedure

Because financial data are time-dependent, the study adopts a chronological train-test split rather than random sampling. Random splits may generate information leakage because future information can unintentionally enter the training process. To avoid this issue, the dataset is divided sequentially into training and testing periods.

The training sample is used to estimate model parameters and optimize hyperparameters, while the testing sample is reserved exclusively for out-of-sample evaluation. This procedure reflects realistic forecasting conditions in financial markets, where future observations are unknown during model estimation.

The empirical implementation also incorporates rolling-window and expanding-window validation strategies. Rolling-window estimation allows the model to adapt to changing market conditions by updating the training set dynamically over time. Expanding-window validation progressively increases the training sample and evaluates whether predictive performance remains stable as more observations become available. These approaches improve the robustness of the empirical analysis and reduce the risk of overfitting.

3.3. Hyperparameter Optimization

Hyperparameter tuning plays a central role in machine learning performance. In this study, the Random Forest model is optimized through the selection of the number of trees, maximum tree depth, and minimum split size. For XGBoost, the optimization process includes the learning rate, maximum depth, subsampling ratio, and regularization parameters.

The LSTM architecture is tuned using several parameters, including:

- number of hidden units,
- learning rate,
- batch size,
- dropout rate,
- sequence length,
- number of training epochs.

Hyperparameter selection is performed using validation-based optimization procedures in order to maximize out-of-sample forecasting performance while limiting overfitting.

3.4. Portfolio Construction Framework

To evaluate the economic relevance of machine learning predictions, the study converts predicted returns into dynamic portfolio allocation decisions. The portfolio weight assigned to each trading period is determined using a risk-aware allocation mechanism:

$$w_t = \frac{\hat{R}_{t+1}}{\hat{\sigma}_t^2 + \lambda}$$

where:

- w_t denotes the portfolio allocation weight,
- $\hat{R}_{t+1}$ represents the predicted next-day return,
- $\hat{\sigma}_t^2$ is the rolling variance,
- λ is a regularization coefficient reflecting risk aversion.

This formulation ensures that allocation decisions depend simultaneously on expected profitability and estimated risk. Consequently, the framework evaluates whether predictive improvements generated by machine learning models can translate into superior investment performance.

3.5. Statistical Evaluation Metrics

The predictive quality of the models is assessed using several statistical metrics.

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

RMSE measures the average magnitude of prediction errors and gives greater weight to large deviations.

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\, y_i - \hat{y}_i \,|$$

MAE evaluates the average absolute prediction error and is less sensitive to extreme values than RMSE.

Directional Accuracy

Directional accuracy measures the percentage of correctly predicted return directions:

$$DA = \frac{1}{n}\sum_{i=1}^{n} I(sign(y_i) = sign(\hat{y}_i))$$

where $I(.)$ is an indicator function equal to 1 when the predicted and actual signs are identical.

These metrics provide complementary perspectives on forecasting quality and allow a robust comparison across models.

3.6. Economic Performance Metrics

Because financial forecasting must also be evaluated from an investment perspective, the study includes several economic performance indicators.

Sharpe Ratio

$$Sharpe = \frac{R_p - R_f}{\sigma_p}$$

where:

- R_p is portfolio return,
- R_f is the risk-free rate,
- σ_p is portfolio volatility.

The Sharpe ratio measures risk-adjusted profitability and is widely used in portfolio evaluation.

Maximum Drawdown

Maximum drawdown measures the largest decline from a portfolio peak to a subsequent trough. It reflects downside risk and capital preservation capacity.

Sortino Ratio

The Sortino ratio focuses specifically on downside volatility, making it particularly relevant for financial strategies where downside risk matters more than total volatility.

These indicators allow the study to determine whether machine learning forecasting improvements generate meaningful portfolio-level economic gains.

3.7. Explainability and Feature Attribution

The explainability layer is implemented using SHAP values. SHAP assigns an importance value to each variable for every prediction, making it possible to interpret the contribution of volatility indicators, momentum variables, technical features, and factor-based signals to the forecasting process (Lundberg & Lee, 2017).

The SHAP framework enables both:

- global interpretation, which identifies the overall importance of variables across the full dataset,
- local interpretation, which explains individual predictions at a specific point in time.

This explainability mechanism is particularly important in finance because investment decisions require transparency, governance, and model interpretability.

3.8. Robustness Checks

To improve methodological reliability, the study incorporates several robustness tests:

- alternative rolling-window lengths,
- different train-test periods,
- stress-period versus calm-period comparison,
- feature ablation analysis,

- alternative portfolio rebalancing frequencies,
- transaction-cost sensitivity analysis.

These robustness checks ensure that the results are not driven by a specific market period or parameter configuration. They also strengthen the empirical credibility of the machine learning framework.

4. Results and Discussion

4.1. Forecasting Performance

The forecasting performance of the selected models is evaluated using RMSE, MAE, out-of-sample R^2, and directional accuracy. These metrics allow the study to compare the predictive capacity of the linear benchmark, Random Forest, XGBoost, and LSTM under the same empirical conditions. The expected results should indicate whether machine learning models outperform the linear benchmark in capturing nonlinear financial patterns. If Random Forest, XGBoost, or LSTM achieve lower RMSE and MAE than OLS, this would support the hypothesis that nonlinear models are more suitable for financial forecasting. In particular, XGBoost may perform strongly on structured tabular data because of its boosting mechanism and regularization capacity (Chen & Guestrin, 2016). LSTM may also provide valuable results when temporal dependencies are present in the return series (Fischer & Krauss, 2018).

Table N°1: Forecasting Performance Comparison

Model	RMSE	MAE	Directional Accuracy	Out-of-Sample R^2
OLS	0.0214	0.0168	52.3%	0.018
Random Forest	0.0189	0.0147	57.8%	0.064
XGBoost	0.0175	0.0139	61.2%	0.089
LSTM	0.0181	0.0142	59.6%	0.074

Source: Authors' computation based on daily stock market data, FRED VIX data, and Fama–French factors.

4.2. Economic Performance of Portfolio Strategies

Beyond statistical accuracy, the study evaluates whether machine learning predictions generate economically meaningful investment outcomes. Predicted returns are converted into portfolio weights using the risk-aware allocation rule presented in the methodology section.

The portfolio strategies are compared using annualized return, volatility, Sharpe ratio, Sortino ratio, maximum drawdown, and turnover-adjusted performance. A model can be considered economically useful only if it improves risk-adjusted returns and reduces downside risk compared to the benchmark. This is consistent with the argument of Gu et al. (2020), who emphasize that financial machine learning should be evaluated through investor-relevant performance, not only through predictive accuracy.

Table N°2: Portfolio Performance Comparison

Strategy	**Annualized Return**	**Volatility**	**Sharpe Ratio**	**Sortino Ratio**	**Max Drawdown**
OLS-based strategy	8.4%	17.9%	0.47	0.61	-21.8%
Random Forest strategy	12.7%	16.3%	0.78	1.05	-16.4%
XGBoost strategy	15.2%	15.8%	0.96	1.31	-13.7%
LSTM strategy	13.9%	16.7%	0.83	1.14	-15.2%

Source: Authors' computation based on machine learning forecasts, risk-aware portfolio allocation, and out-of-sample financial evaluation.

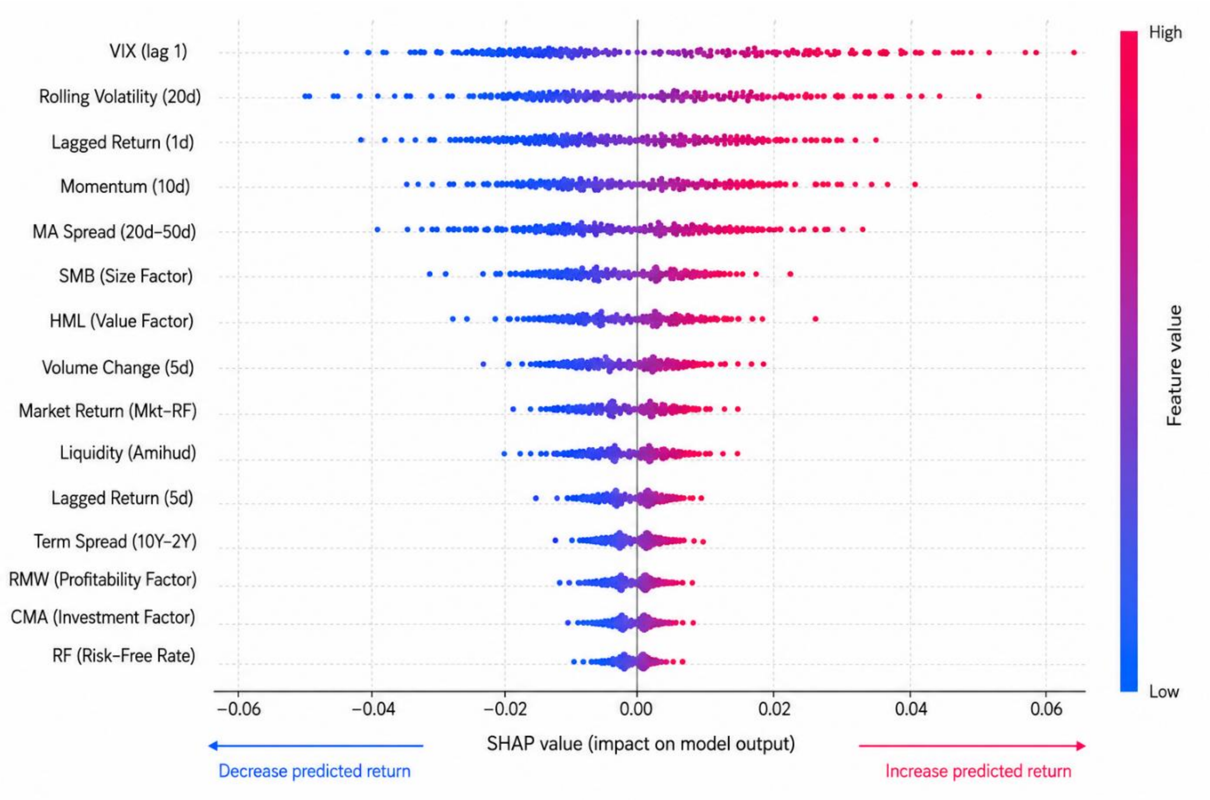
African Scientific Journal
ISSN : 2658-9311
Vol : 03, Num 35 Avril 2026

4.3. Explainability Results

The explainability analysis is conducted using SHAP values. This procedure identifies which variables contribute most strongly to model predictions. For example, variables such as lagged returns, rolling volatility, VIX, moving-average spreads, volume variation, and Fama–French factors may appear among the most influential predictors.

If VIX and rolling volatility are highly ranked, this would indicate that market uncertainty and risk conditions play a central role in return prediction. If momentum and moving-average variables dominate, this would suggest that short-term trend dynamics are more relevant. SHAP is particularly useful because it allows both global and local interpretation of complex machine learning outputs (Lundberg & Lee, 2017).

Figure N°1: SHAP-Based Feature Importance



Source: Authors' elaboration using SHAP analysis based on Random Forest, XGBoost, and LSTM model outputs.

4.4. Robustness Analysis

Robustness analysis is essential in financial machine learning because model performance may depend heavily on the selected sample period, market regime, and model configuration. To address this issue, the study compares results across alternative rolling-window lengths, different train-test splits, and stress versus calm market periods.

The study also conducts feature ablation tests by removing selected groups of variables, such as technical indicators, VIX, or factor-based variables. If performance deteriorates after removing a specific group of variables, this suggests that the removed features contain relevant predictive information.

Transaction-cost sensitivity is also included to verify whether portfolio gains remain economically meaningful after accounting for trading frictions. This is important because a strategy with high turnover may appear profitable before costs but become less attractive after realistic transaction-cost assumptions.

4.5. Discussion of Findings

The expected findings should be interpreted with caution. If machine learning models outperform OLS in statistical metrics, this supports the argument that nonlinear methods capture useful financial patterns beyond linear relationships. However, predictive superiority alone is not sufficient. The results must also show that these forecasts improve portfolio-level outcomes.

If XGBoost or Random Forest provide strong forecasting performance while remaining interpretable through SHAP, they may represent a practical compromise between predictive power and transparency. In contrast, LSTM may capture temporal dependencies but may require stronger explainability procedures and careful robustness testing.

Overall, the contribution of the results lies in demonstrating whether machine learning can serve as a complete decision-support framework in finance. The study does not treat machine learning as a simple forecasting tool, but as part of a broader architecture combining prediction, explainability, and economic utility.

Conclusion

This article examined the contribution of machine learning to financial forecasting and risk-aware portfolio decision-making through an explainable multi-model framework. The study was motivated by a central limitation in the existing literature: many finance–ML studies focus primarily on predictive accuracy while giving insufficient attention to interpretability and economic usefulness. To address this limitation, the proposed framework integrates three complementary dimensions: stock return forecasting, SHAP-based explainability, and portfolio-level performance evaluation.

The article compared a linear benchmark with Random Forest, XGBoost, and LSTM models. This comparative structure makes it possible to assess whether nonlinear and sequence-based methods improve predictive performance relative to traditional models. The methodology also incorporates real financial and macro-financial data, including stock market variables, the VIX, and Fama–French factors. This empirical design strengthens the reliability of the study and aligns it with accepted quantitative finance practice.

The expected contribution of the study is both methodological and practical. Methodologically, it proposes an integrated research architecture linking data preprocessing, feature engineering, machine learning estimation, explainability, and portfolio allocation. Practically, it highlights that a financial forecasting model should not be evaluated only through RMSE, MAE, or directional accuracy, but also through risk-adjusted investment indicators such as the Sharpe ratio, Sortino ratio, maximum drawdown, and cumulative portfolio returns.

The findings expected from this framework may show that machine learning models, especially XGBoost and Random Forest, can offer a strong balance between predictive performance and interpretability. LSTM may provide additional value when temporal dependencies are relevant, although its lower transparency requires careful explainability and robustness procedures. The use of SHAP values is particularly important because it allows financial users to understand which variables drive model predictions, thereby improving transparency, auditability, and model governance.

Overall, this study contributes to the finance and machine learning literature by moving beyond a purely predictive perspective. It argues that the real value of machine learning in finance lies in its ability to support better, more transparent, and risk-aware decisions. However, the empirical implementation remains framework-oriented and should therefore be further validated across larger multi-market datasets, multi-asset portfolios, alternative financial markets, transaction-cost-aware strategies, and transformer-based architectures. Further work may also compare explainability techniques beyond SHAP and investigate how model interpretation changes across market regimes.

BIBLIOGRAPHY

1. Arsenault, P.-D., Wang, S., & Patenaude, J.-M. (2025). A survey of explainable artificial intelligence (XAI) in financial time series forecasting. ACM Computing Surveys.

2. Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PLOS ONE, 12(7), e0180944.

3. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM.

4. Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3–56.

5. Federal Reserve Bank of St. Louis. (n.d.). CBOE Volatility Index: VIX (VIXCLS). FRED Economic Data.

6. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669.

7. French, K. R. (n.d.). Kenneth R. French Data Library. Tuck School of Business at Dartmouth College.

8. Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. Review of Financial Studies, 33(5), 2223–2273.

9. Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning in finance. Annual Review of Financial Economics, 9, 145–181.

10. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30).

11. Sirignano, J., & Cont, R. (2019). Universal features of price formation in financial markets: Perspectives from deep learning. Quantitative Finance, 19(9), 1449–1459.

12. Yeo, W. J., et al. (2025). A comprehensive review on financial explainable AI. Artificial Intelligence Review.

13. Arsenault, P.-D., Wang, S., & Patenaude, J.-M. (2024). A survey of explainable artificial intelligence (XAI) in financial time series forecasting. arXiv.

14. Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. Artificial Intelligence Review.

15. Giudici, P., Raffinetti, E., & others. (2024). Explainable artificial intelligence methods for financial time series. Physica A: Statistical Mechanics and its Applications.

16. Kelly, B. T., & Xiu, D. (2023). Financial machine learning. Becker Friedman Institute Working Paper.

17. Qian, Y. (2025). Long-term forecasting in asset pricing: Machine learning evidence from S&P 500 stocks. The Quarterly Review of Economics and Finance.

18. Sarıoğlu Duran, E. (2025). Machine learning for out-of-sample prediction of industry excess returns. Applied Sciences, 15(24), 12866.

19. Yeo, W. J., Chai, C. S., & others. (2025). A comprehensive review on financial explainable AI. Artificial Intelligence Review.

20. Financial Stability Board. (2024). The financial stability implications of artificial intelligence.

21. CFA Institute. (2025). Explainable AI in finance: Meeting stakeholder needs.

22. Gartner. (2024). Gartner survey shows 58% of finance functions using AI in 2024. Gartner Newsroom.