Machine learning-driven stock price prediction for enhanced investment strategy

Omaima Guennioui, Dalila Chiadmi, Mustapha Amghar
Mohammadia School of Engineering, Mohammed V University in Rabat, Rabat, Morocco

Article Info

Article history:

Received Apr 7, 2024
Revised Jul 4, 2024
Accepted Jul 9, 2024

Keywords:

Global forecasting models
Investment strategy
Machine learning
Portfolio optimization
Stock forecasting
Stock market

ABSTRACT

Forecasting stock prices, a task complicated by the inherent volatility of the stock market, poses a significant challenge. The ability to accurately forecast stock prices is crucial, as it provides investors with crucial insights, enabling them to make informed strategic decisions. In this paper, we propose a novel investment strategy that relies on predicting stock prices. Our approach utilizes a hybrid predictive model that combines light gradient-boosting machine (LightGBM) and extreme gradient boosting (XGBoost). This model is designed to generate short to medium-term forecasts for a wide range of stocks. The strategy has shown promising results, surpassing the local market indices used as benchmarks in terms of both risk and return. Our findings demonstrate the strategy's effectiveness in both upward and downward market trends, underscoring its potential as a robust tool for portfolio management in diverse market conditions.

Corresponding Author:

Omaima Guennioui
Mohammadia School of Engineering, Mohammed V University in Rabat
Rabat, Morocco
Email: omaimaguennioui@research.emi.ac.ma

1. INTRODUCTION

Stock market forecasting is a complex challenge, the semi-strong form of the market efficiency hypothesis [1], which is broadly recognized in finance, suggests that security prices rapidly respond to new information available to the public. Additionally, market data are highly volatile and filled with noise, making stock price prediction a particularly arduous task. Despite this, numerous markets anomalies exist that challenge the efficient market theory [2], highlighting the importance of developing trading strategies that can effectively trade stocks and outperform the market.

In addressing the challenge of predicting financial markets, especially with the vast amount of data on stocks, machine learning (ML) shows great promise. Studies have applied machine learning and deep learning techniques to try to predict stock market movements [3]–[5]. In this context, Hu *et al.* [6] highlighted the effectiveness of light gradient-boosting machine (LightGBM) for medium to long-term forecasting, selected based on a comprehensive evaluation of various machine learning models using different performance metrics. Guo *et al.* [7] developed an efficient long short-term memory-LightGBM hybrid model for stock price prediction, and Yun *et al.* [8] enhanced predictive accuracy with their hybrid genetic extreme gradient boosting (GA-XGBoost) system through optimal feature selection. While these studies demonstrate machine learning's capability in predicting stock prices, they often do not apply these insights to form trading or investment strategies, a finding also noted in the study by Sonkavde *et al.* [9]. In contrast, some research has successfully combined stock price prediction with strategy development. Picasso *et al.* [10] used a neural network for market trend classification, demonstrating its practicality in high-frequency trading simulations. Li *et al.* [11] explored the effects of various indicators on stock returns, with LightGBM emerging as the most effective among the algorithms tested, leading to its integration into a profitable investment strategy. Although there has been significant advancement in financial forecasting, especially in stock price prediction, the application of these forecasts in conceiving trading strategies remains relatively explored. Moreover, recent advancements have been promising in the field of global forecasting models, which are trained on multiple time series data [12]–[14]. These models, focusing on cross-series data, aim to efficiently manage complexity and prevent overfitting on a larger scale [15]–[17].

Recognizing these challenges and opportunities, this paper introduces a novel investment strategy that effectively utilizes predictive insights from advanced machine learning models. This strategy integrates a hybrid model composed of LightGBM and XGBoost, specifically designed to predict short and medium-term future prices across a diverse range of stocks. Such predictive capabilities are crucial for informed stock selection and strategic investment decision-making. The strategy is structured around two key components: the initial focus is on accurately forecasting future stock prices, and the second component develops decision rules that leverage these forecasts in the buying and selling of stocks, particularly applying this method to the relatively unexplored Moroccan market. This approach seeks to discover potential growth opportunities and adeptly manage the unique challenges presented by this emerging market. Moreover, while there have been significant advancements in the field of financial forecasting using machine learning, the direct application of these models in developing comprehensive investment strategies, especially in emerging markets, has not been extensively explored. This study aims to bridge this gap by demonstrating a practical framework for integrating advanced forecasting models into effective investment strategies.

2. METHOD

2.1. Prediction component

In this study, we introduce the global LightGBM-XGBoost hybrid model, employing a global approach to predict stock prices by aggregating results from two individual models: LightGBM and XGBoost. On one hand, LightGBM, a machine learning algorithm introduced by Microsoft in 2017 and part of the gradient boosting decision tree family [18], is optimized for large-scale supervised learning tasks such as regression, classification, and ranking. It achieves high accuracy through an iterative process of building decision trees and employing a leaf-wise split strategy, optimizing each tree leaf's feature value to enhance overall accuracy gains. On the other hand, XGBoost, introduced by Chen and Guestrin [19], is a widely-used open-source software library that implements gradient boosting machine algorithms. It improves upon the conventional gradient boosted regression trees (GBRT) model by sequentially building basic regression trees, with each new model attempting to correct the errors of its predecessor. XGBoost employs a level-wise approach for tree growth, dividing all nodes at the same level before advancing, resulting in a more balanced tree structure and preventing overfitting. Renowned for its high problem-solving performance, XGBoost, along with LightGBM, stands as a popular choice within the data science community for its versatility in handling diverse structured and unstructured datasets and delivering superior predictive accuracy compared to other machine learning algorithms [20], [21]. This is particularly relevant in the context of the global approach which addresses the challenges posed by local models. Local models often become more complex as the dataset grows, surpassing the constant complexity of a global model built on the same data. Calibrating a local model for a single series might be swift, but adapting multiple models to a series set is more time-consuming. Additionally, local error generalization limits often exceed those of global errors [22].

Global LightGBM-XGBoost utilizes ensemble learning, which combines the strengths of multiple independent models to improve performance and minimize errors. We specifically used a voting ensemble method, integrating predictions from both the LightGBM and XGBoost models, which were trained separately on the same dataset. Since both models showed similar accuracy, we opted for the average voting method. This technique averages the predictions from each model at every data point, thus blending their strengths to enhance the final output. Averaging helps neutralize biases or variances specific to any single model, making our predictions more robust and reliable. This approach is particularly effective when models agree on general trends but differ slightly in their detailed forecasts, resulting in a balanced output. Recognizing the significant influence of hyperparameters on the performance of machine learning models, as demonstrated by previous research [23], [24] selecting optimal hyperparameters is essential. We employed the grid search method to identify the best hyperparameter combinations for our dataset. Grid search methodically evaluates various combinations within a specified range to determine which configuration achieves the highest model accuracy [25], [26]. This comprehensive approach ensures the identification and utilization of the most effective settings, thereby optimizing model performance [27]. Additionally, we integrated cross-validation with grid search in our study to enhance the generalization capability of the XGBoost and LightGBM models, crucial for accurate stock price predictions.

Global LightGBM-XGBoost was designed to make predictions across multiple time horizons, specifically at intervals of 1, 5, and 21 days, corresponding respectively to predictions for the next day, the following week, and an approximation of a month. This reflects the fact that stocks are traded on business days only. Figure 1 illustrates the workflow diagram of the global LightGBM-XGBoost model. The process begins with data preprocessing, which involves the normalization of data. Both the LightGBM and XGBoost models are initially configured with hyperparameters that are subsequently fine-tuned to optimize performance. The optimized models are then combined into a single ensemble model, referred to as global LightGBM-XGBoost.

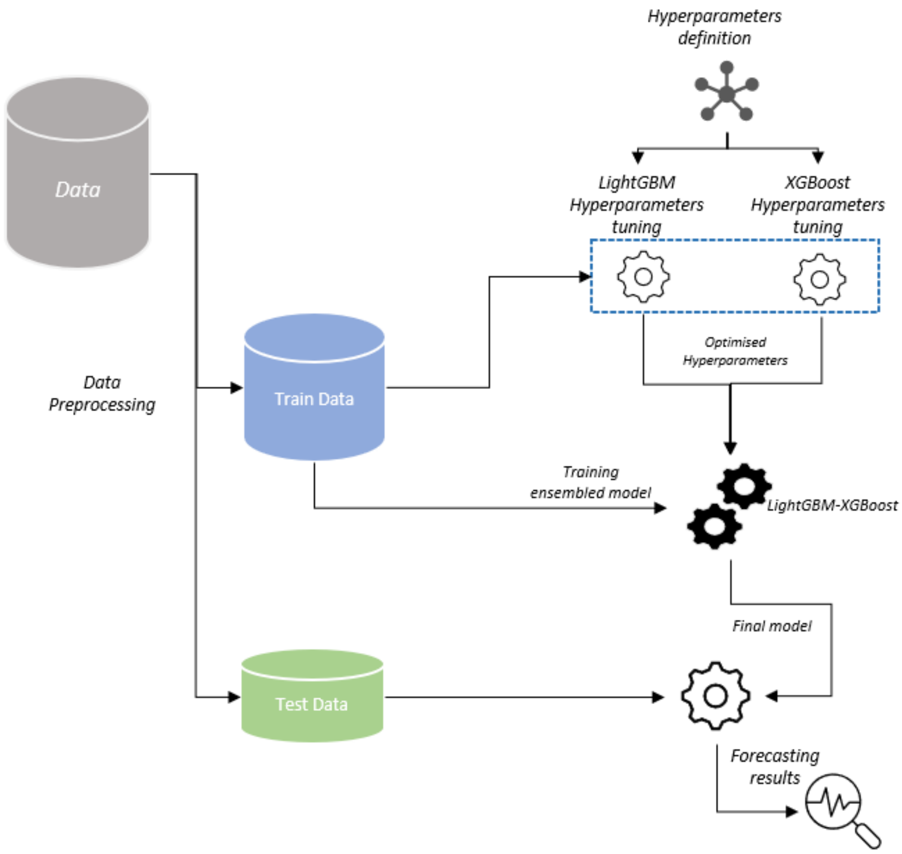


Figure 1. Workflow diagram for Global LightGBM-XGBoost predictive model

2.2. Proposed strategy

In this paper, we examine an investment portfolio comprising n stocks, rebalanced at the start of each month. Our method involves key steps as follows:

- Prediction phase: At the end of each month, we forecast the movement of these n stocks for the following day, week, and month. This forecasting enables us to estimate the adjusted expected returns (AER) of each stock for the upcoming month, as calculated in (1).
- Stock ranking and rebalancing: The stocks are ranked based on their AER, and on the first business day of each month, we sell a percentage of the holdings in the two stocks projected to be the least promising, and simultaneously purchase an equivalent amount of the two stocks predicted to be the most profitable. This process forms a part of our active portfolio management strategy. The decision to focus on just two stocks in our study is driven by a desire to maintain a restrictive approach in identifying investment opportunities, thereby effectively managing risk. By limiting the number of stocks, we actively trade, we not only control our exposure to market volatility but also minimize transaction costs, enhancing the overall efficiency and potential profitability of our investment strategy.
- Daily portfolio valuation: We then track the performance of our implemented stock portfolio on a daily basis to monitor its performance.

The formula for the monthly asset evolution estimation, which is central to our strategy, is given by (1):

$$AER_{i,t} = W_1 * P_{i,(t,t+1)} + W_5 * P_{i,(t,t+5)} + W_{21} * P_{i,(t,t+21)} \quad (1)$$

where

$$P_{i,(t,t+s)} = \frac{S_{i,t+s} - S_{i,t}}{S_{i,t}} \quad (2)$$

In the formula, $S_{i,t}$ represents the closing price of stock i, where i ranges from 1 to n, on day t, with $\{W_1, W_5, W_{21}\}$ being the weights for each horizon.

We primarily use the root mean square error (RMSE) to determine the weights. RMSE penalizes larger errors more than smaller ones, thereby enhancing the accuracy of our analysis. By giving greater weight to significant errors, we minimize the risk of substantial distortions in the calculated adjusted expected returns, thereby ensuring more reliable estimations. We assign weights $\{W_1, W_5, W_{21}\}$ inversely proportional to the RMSE for each prediction horizon, meaning higher errors result in lower weights.

Based on these computed AERs, we rank the stocks, identifying the two with the lowest AER for the upcoming period and those with the highest AER. We then sell a percentage of the holdings in the two stocks projected to be the least promising, and simultaneously purchase an equivalent amount of the two stocks predicted to be the most profitable. This strategy allows us to reduce our investment in less promising stocks and reinvest in those with higher predicted returns, thereby optimizing our portfolio for future performance.

2.3. Dataset description

Data for this study were gathered from the Casablanca stock exchange website [28], covering the period from January 2012 to January 2023. In the Moroccan stock market, over half of the market capitalization of the Moroccan all shares index (MASI)-the index that reflects the overall performance of stocks listed on the Casablanca stock exchange is concentrated in three primary sectors. These sectors include Banking, a pivotal element of Morocco's economy; building and construction materials, indicative of the country's expanding construction industry; and Telecommunication, dominated by Itissalat Al-Maghrib, Morocco's leading telecom operator with extensive operations in approximately ten francophone African countries. Our study will subsequently focus specifically on these three crucial sectors, the composition of which is detailed in Table 1.

Table 1. Composition of studied sectors

Sector	Ticker	Stock
Banking	BOA	BANK OF AFRICA
	CIH	CIH
	CDM	CDM
	BCI	BMCI
	BCP	BCP
	ATW	ATTIJARIWAFA BANK
Building and	JET	JET CONTRACTORS
Construction materials	AFI	AFRIC INDUSTRIES SA
	SID	SONASID
	TGC	TGCC S.A
	LHM	LAFARGEHOLCIM MAROC
	CMA	CIMENTS DU MAROC
	ALM	ALUMINIUM DU MAROC
	COL	COLORADO
Telecommunication	IAM	ITISSALAT AL-MAGHRIB

We collected daily opening and closing stock prices for the stocks listed in Table 1 throughout the study period, ensuring alignment across the stocks. For our analysis, we assumed that all buy and sell orders would be executed at the opening price on the first day of each month. To develop our predictive model, we divided the raw dataset into training and testing sets, with 80% allocated for training and 20% for testing. Note that our study excluded companies listed after 2012, such as TGCC S.A (building and construction materials) as there was insufficient data available.

Transaction costs were also taken into account. Typically, trade fees comprise broker fees, exchange fees and regulatory fees. In real-world trading environments, a fund or trading firm might incur varying execution costs for several reasons. However, for our study, we standardized the transaction cost to 35 basis points of the trade's value after considering various scenarios, which we think is a reasonable estimate for our purposes.

2.4. Evaluation metrics
2.4.1. Evaluation of prediction accuracy
We assess the accuracy of our global LightGBM-XGBoost model using established metrics such as the root mean square error (RMSE), mean absolute error (MAE), and R2 score (R2), calculated as (3)-(5):

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_j - \tilde{y}_j)^2} \quad (3)$$

$$MAE = \frac{1}{N}\sum_{j=1}^{N}|y_j - \tilde{y}_j| \quad (4)$$

$$R2 = 1 - \frac{\sum_{j=1}^{N}(y_j - \tilde{y}_j)^2}{\sum_{j=1}^{N}(y_j - \bar{y}_j)^2} \quad (5)$$

In this context, the symbol y denotes the actual closing price of the test set, while the symbol $\tilde{y}$ represents the predicted value of that price. The symbol $\bar{y}$ represents the mean value of y and N denotes the number of data points in the test set.

2.4.2. Investment strategy assessment
The proposed investment strategy is evaluated using overall performance comparison and risk-adjusted metrics. Portfolio return is calculated using (6).

$$Rp = \frac{V_f - V_i}{V_i} \quad (6)$$

where, V_f and V_i represent the final and initial values of the portfolio, respectively.

For risk-adjusted return evaluation, we use the Sharpe ratio, which measures the average excess return earned for each unit of risk above the risk-free rate. A higher Sharpe ratio indicates a more attractive risk-adjusted return, while a negative Sharpe ratio points to a risk-free asset outperforming. The Sharpe ratio formula is:

$$Sharpe\ Ratio = \frac{Rp - Rf}{\sigma} \quad (7)$$

Here, Rf is the risk-free rate, set at 1.6% based on the Moroccan 1-year bond quotation during the test period, and σ is the portfolio volatility.

Furthermore, we utilize Jensen’s alpha to measure absolute performance. It represents the annualized return of the portfolio, adjusted for the risk-free yield, minus the benchmark return multiplied by the stock’s beta during the same period. Jensen’s alpha indicates the excess return from deviating from the benchmark [29]. Its formula is:

$$Jensen's\ alpha = Rp - (Rf + (Rm - Rf) * \beta) \quad (8)$$

Lastly, we calculate the beta (β), a measure of the portfolio's volatility relative to the overall market. Beta is calculated as (9):

$$\beta = \frac{Cov\,(Rp - Rf, Rm - Rf)}{Var(\,Rm - Rf)} \quad (9)$$

where Rm represents market return.

3. RESULTS AND DISCUSSION
To enhance the prediction results and mitigate issues which often occur due to manual parameter specification, parameter tuning is essential. In our research, we utilized GridSearchCV for this purpose. The outcomes of the parameter tuning process for the LightGBM and XGBoost models are detailed in Tables 2 and 3, respectively.

The hyperparameter range chosen for our model across different prediction horizons in our study were selected to align with the specific requirements of each timeframe. This calibration ensures that our model is tuned to accurately capture market trends and dynamics, whether for short or medium-term forecasting. Table 4 showcases the performance results of the global LightGBM-XGBoost model and its comparison with the LightGBM and XGBoost models, all utilizing optimized parameters across the studied time horizons. The models were evaluated based on metrics such as RMSE, MAE, and R2 score.

Table 2. Hyperparameter tuning results for LightGBM

Horizon	Hyperparameter	Description	Optimal value
1 day	n estimators	Number of iterations	300
	learning rate	Controls the contribution of each tree in the ensemble	0.02
	bagging fraction	Percentage of data used in each iteration	0.3
	reg lambda	Regularization on weights to avoid overfitting	0.3
	reg alpha	L1 regularization	0.3
5 days	n estimators	Number of iterations	200
	learning rate	Controls the contribution of each tree in the ensemble	0.02
	bagging fraction	Percentage of data used in each iteration	0.6
	reg lambda	Regularization on weights to avoid overfitting	0.1
	reg alpha	L1 regularization	1.0
21 days	n estimators	Number of iterations	1000
	learning rate	Controls the contribution of each tree in the ensemble	0.01
	bagging fraction	Percentage of data used in each iteration	0.7
	reg lambda	Regularization on weights to avoid overfitting	1.0
	reg alpha	L1 regularization	1.0
	max_depth	Controls the depth of the trees	7

Table 3. Hyperparameter tuning results for XGBoost

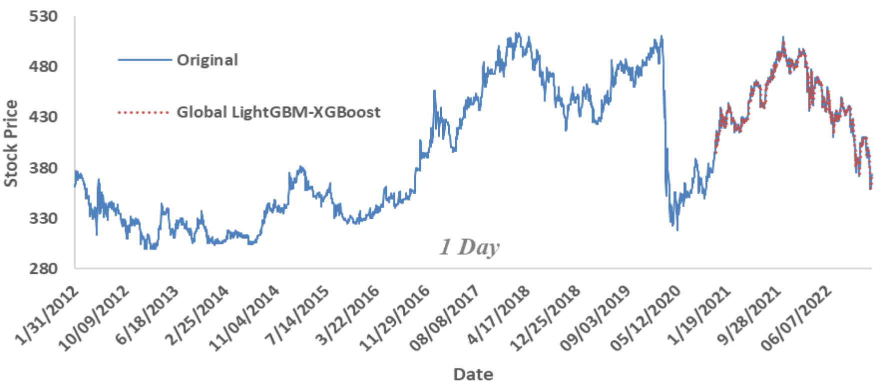
Horizon	Hyperparameter	Description	Optimal value
1 day	n estimators	Number of iterations	300
	learning rate	Controls the contribution of each tree in the ensemble	0.02
	reg lambda	Regularization on weights to avoid overfitting	0.1
	subsample	Sample fraction used to train each tree	0.6
5 days	n estimators	Number of iterations	300
	learning rate	Controls the contribution of each tree in the ensemble	0.02
	reg lambda	Regularization on weights to avoid overfitting	0.01
	subsample	Sample fraction used to train each tree	0.3
21 days	n estimators	Number of iterations	200
	learning rate	Controls the contribution of each tree in the ensemble	0.02
	reg lambda	Regularization on weights to avoid overfitting	0.1
	subsample	Sample fraction used to train each tree	0.9
	max_depth	Controls the depth of the trees	3

Table 4. Forecasting performance by timeframe

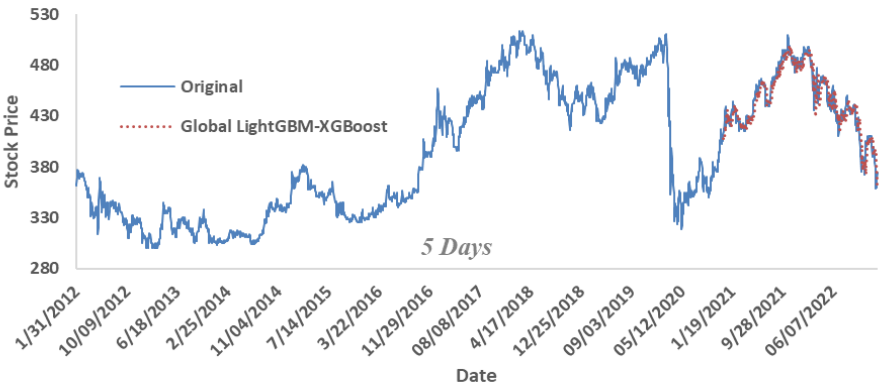
Horizon	Method	RMSE	MAE	R2 score
1 day	LightGBM	0.0194	0.0121	0.991
	XGBoost	0.0196	0.0122	0.990
	Global LightGBM-XGBoost	0.0192	0.0119	0.991
5 days	LightGBM	0.0370	0.0256	0.966
	XGBoost	0.0371	0.0255	0.967
	Global LightGBM-XGBoost	0.0367	0.0253	0.967
21 days	LightGBM	0.0715	0.0531	0.871
	XGBoost	0.0719	0.0532	0.871
	Global LightGBM-XGBoost	0.0713	0.0529	0.872

The RMSE and MAE values produced by our model for the 1-day horizon, at 0.0192 and 0.0119 respectively, demonstrate a significant degree of accuracy in short-term predictions. This highlights the model's capability to effectively capture immediate market movements with minimal error. Additionally, the high R2 score of 0.991 further reinforces the model's strong fit and predictive power within this timeframe. As the forecasting horizon extends to 5 and 21 days, we observe an incremental increase in RMSE and MAE values. This pattern is typical in longer-term financial forecasting, where market uncertainty and volatility present greater challenges. However, the model still maintains a high level of accuracy, as evidenced by R2 scores of 0.967 and 0.872 for the 5-day and 21-day periods, respectively. In comparing our Global LightGBM-XGBoost model with the individual LightGBM and XGBoost models, it consistently outperforms them across all timeframes. The combined model shows marginally lower RMSE and MAE values for the 1-day horizon, highlighting its superior accuracy. Even as errors naturally increase over the 5-day and 21-day horizons, the global LightGBM-XGBoost model maintains tighter error margins and exhibits greater predictive accuracy. The strong performance of this model supports our decision to use it as the main part in developing our investment strategy. Further enhancement might be explored through advanced big data frameworks such as those discussed by Yousfi *et al.* [30], which offer additional methodologies for refining data analysis.

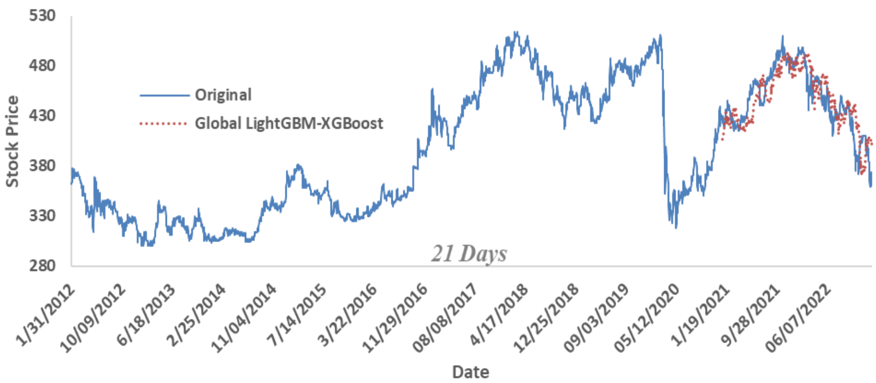
The observed shift in precision across different horizons prompted us to incorporate different weights for the adjusted expected returns (AER) calculation to reflect the varying levels of prediction accuracy at each horizon. This weighted approach allows for a more nuanced and accurate assessment of expected returns, taking into account the changing reliability of the model's predictions over different timeframes. Figure 2 displays a comparison of the predicted closing prices from the global LightGBM-XGBoost model for the largest market capitalization in MASI, Attijariwafa bank, against the actual values over three-time horizons in Figures 2(a) 1 day, 2(b) 5 days, and 2(c) 21 days.



(a)



(b)



(c)

Figure 2. Comparing the forecasted closing prices of ATTIJARIWAFA BANK with the actual values across three timeframes: (a) 1-day, (b) 5-day, and (c) 21-day horizons

In our study, results from our predictions were utilized in the backtesting of our investment strategy, focusing specifically on the years 2021 and 2022. We evaluated our strategy over these complete years, employing a year-to-date approach. Notably, the selected period is of particular interest: 2021 witnessed a strongly positive trend in the stock market, with the MASI growing by +18%, while 2022 experienced a distinctly negative trend, with a -20% decline in the MASI. This contrasting scenario provides a valuable opportunity to test our strategy under diverse market conditions.

Additionally, we have determined the weights for each prediction horizon in the calculation of the AER based on our prediction results. The derived weights for {W1, W5, W21} are {56%, 29%, 15%}, respectively. By feeding our algorithm with various stock price predictions across different horizons, the algorithm computes the AERs for each date. This process generates a matrix by date, indicating which stocks to sell and buy for the upcoming month, effectively guiding our investment decisions.

The initial capital for our investment portfolio was set at 100 million Moroccan Dirhams. We consider that the portfolio is initially structured with the same weightings as the composite index of the three sectors under study, ensuring a representative and balanced outset. Additionally, we define a coefficient of 5%, which represents the portion of the held volume of stocks to be sold. This means that we plan to divest 5% of the volume held in the two stocks anticipated to decrease in value, and then reinvest the equivalent amount equitably into the stocks projected to rise. The results of our strategy for the years 2021 and 2022 are presented in Tables 5 and 6, respectively.

Table 5. Backtesting results for 2021

	Portfolio strategy	Composite index
Return	13.3%	11.1%
Volatility	7.8%	7.84%
Sharpe ratio	1.39	1.14
Beta	0.97	
Jensen’s alpha	0.022	

Table 6. Backtesting results for 2022

	Portfolio strategy	Composite index
Return	-24.4%	-25.1%
Volatility	13.2%	13.5%
Sharpe ratio	-2.10	-2.13
Beta	0.98	
Jensen’s alpha	0.003	

In 2021, our portfolio strategy outperformed the composite index, delivering a performance of 13.3% compared to 11.1%. This indicates a robust strategy capable of generating higher returns. The Sharpe ratio of 1.39 for our strategy, exceeding the index's 1.14, reflects superior risk-adjusted returns. The beta value close to 1 suggests that our portfolio's volatility is in line with the market, while a positive Jensen's alpha of 0.022 indicates that our strategy achieved additional returns over the expected market returns, given the level of market risk. In 2022, despite a decline in the market, our proposed portfolio strategy showed resilience. It registered a better return, exceeding the composite index by +0.7%. The Sharpe ratio, although negative for both, was better for our strategy than for the index, suggesting marginally better risk management. The positive, though modest, Jensen's alpha indicates our strategy's capability to generate returns above the market average, even after adjusting for risk.

In an effort to enhance the results for the year 2022, we experimented with increasing the coefficient, which determines the percentage of the held stock volume to be sold, from 5% to 30%. This adjustment aimed to enable more pronounced positioning in our trading strategy, essentially allowing us to adopt a more aggressive approach in response to market conditions. The impact of this strategic modification on our portfolio's performance is detailed in Table 7.

Table 7. 2022 backtesting results with adjusted coefficient

	Portfolio strategy	Composite index
Performance	-22.7%	-25.1%
Volatility	12.8%	13.5%
Sharpe ratio	-2.01	-2.13
Beta	0.94	
Jensen’s alpha	0.013	

Compared to the original 2022 results, the updated strategy exhibits a notable improvement in performance, with the portfolio strategy now showing a reduced loss of -22.7% against the -25.1% of the composite index. This indicates a more effective response to market downturns with the increased coefficient. The portfolio's volatility also decreased to 12.8% from the previous 13.2%. The Sharpe ratio saw a slight improvement to -2.01, compared to the earlier -2.10, indicating a better risk-adjusted return. Notably, Jensen's alpha increased to 0.013, signifying that the adjusted strategy yielded better-than-expected returns against the market after accounting for risk. Overall, the comprehensive results showcased for the years 2021 and 2022 underscore the efficacy of our portfolio strategy in navigating both rising and falling market trends. Using the global LightGBM-XGBoost model's predictions to create our strategy has proven effective, showing that our approach is strong and flexible to handle different market situations.

4. CONCLUSION

This paper introduces a new investment strategy based on predicting future stock prices. We propose a hybrid predictive model that integrates LightGBM and XGBoost, utilizing a global approach for implementation. This model has proven to be effective in providing accurate forecasts in the Moroccan stock market for both short-term and medium-term horizons. Regarding investment strategy, our approach facilitates the selection of stocks likely to increase or decrease in value relative to their current price, enabling effective portfolio rebalancing. Demonstrating promising outcomes, our strategy has surpassed market indices in terms of risk and return, effectively navigating both upward and downward market trends.

There are several directions that can be explored following this project. One possibility includes expanding the approach to the entire market and applying it to other markets as well. Additionally, there is an opportunity to refine the prediction model and data analysis to achieve more precise forecasts for medium and long-term horizons. Adapting the investment strategy to incorporate these long-term predictions could lead to a more comprehensive and robust financial decision-making tool.

REFERENCES

[1] E. F. Fama, "Efficient capital markets: a review of theory and empirical work," *The Journal of Finance*, vol. 25, no. 2, May 1970, doi: 10.2307/2325486.

[2] J. Green, J. R. M. Hand, and X. F. Zhang, "The supraview of return predictive signals," *Review of Accounting Studies*, vol. 18, no. 3, pp. 692–730, Sep. 2013, doi: 10.1007/s11142-013-9231-1.

[3] W. Jiang, "Applications of deep learning in stock market prediction: recent progress," *Expert Systems with Applications*, vol. 184, Dec. 2021, doi: 10.1016/j.eswa.2021.115537.

[4] L. N. Mintarya, J. N. M. Halim, C. Angie, S. Achmad, and A. Kurniawan, "Machine learning approaches in stock market prediction: a systematic literature review," *Procedia Computer Science*, vol. 216, pp. 96–102, 2023, doi: 10.1016/j.procs.2022.12.115.

[5] O. Guennioui, D. Chiadmi, and M. Amghar, "Systematic mapping study on applications of deep learning in stock market prediction," *In: Tabaa, M., Badir, H., Bellatreche, L., Boulmakoul, A., Lbath, A., Monteiro, F. (eds) New Technologies, Artificial Intelligence and Smart Data. INTIS 2022 2023. Communications in Computer and Information Science*, Springer, Cham, vol. 1728, pp. 20–31, 2024, doi: 10.1007/978-3-031-47366-1_2

[6] Q. Hu, S. Qin, and S. Zhang, "Comparison of stock price prediction based on different machine learning approaches," Proceedings of the 2022 International Conference on Bigdata Blockchain and Economy Management (ICBBEM 2022), 2022, pp. 215–231, doi: 10.2991/978-94-6463-030-5_24.

[7] Y. Guo, Y. Li, and Y. Xu, "Study on the application of LSTM-LightGBM model in stock rise and fall prediction," *MATEC Web of Conferences*, vol. 336, Feb. 2021, doi: 10.1051/matecconf/202133605011.

[8] K. K. Yun, S. W. Yoon, and D. Won, "Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process," *Expert Systems with Applications*, vol. 186, p. 115716, Dec. 2021, doi: 10.1016/j.eswa.2021.115716.

[9] G. Sonkavde, D. S. Dharrao, A. M. Bongale, S. T. Deokate, D. Doreswamy, and S. K. Bhat, "Forecasting stock market prices using machine learning and deep learning models: a systematic review, performance analysis and discussion of implications," *International Journal of Financial Studies*, vol. 11, no. 3, Jul. 2023, doi: 10.3390/ijfs11030094.

[10] A. Picasso, S. Merello, Y. Ma, L. Oneto, and E. Cambria, "Technical analysis and sentiment embeddings for market trend prediction," *Expert Systems with Applications*, vol. 135, pp. 60–70, Nov. 2019, doi: 10.1016/j.eswa.2019.06.014.

[11] Z. Li, W. Xu, and A. Li, "Research on multi factor stock selection model based on LightGBM and Bayesian optimization," *Procedia Computer Science*, vol. 214, pp. 1234–1240, 2022, doi: 10.1016/j.procs.2022.11.301.

[12] H. Hewamalage, C. Bergmeir, and K. Bandara, "Global models for time series forecasting: a simulation study," *Pattern Recognition*, vol. 124, Apr. 2022, doi: 10.1016/j.patcog.2021.108441.

[13] C. S. Bojer and J. P. Meldgaard, "Kaggle forecasting competitions: an overlooked learning opportunity," *International Journal of Forecasting*, vol. 37, no. 2, pp. 587–603, Apr. 2021, doi: 10.1016/j.ijforecast.2020.07.007.

[14] T. Januschowski *et al.*, "Criteria for classifying forecasting methods," *International Journal of Forecasting*, vol. 36, no. 1, pp. 167–177, Jan. 2020, doi: 10.1016/j.ijforecast.2019.05.008.

[15] P. Montero-Manso and R. J. Hyndman, "Principles and algorithms for forecasting groups of time series: Locality and globality," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1632–1653, Oct. 2021, doi: 10.1016/j.ijforecast.2021.03.004.

[16] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "M5 accuracy competition: results, findings, and conclusions," *International Journal of Forecasting*, vol. 38, no. 4, pp. 1346–1364, Oct. 2022, doi: 10.1016/j.ijforecast.2021.11.013.

[17] J. Sousa and R. Henriques, "Intersecting reinforcement learning and deep factor methods for optimizing locality and globality in forecasting: a review," *Engineering Applications of Artificial Intelligence*, vol. 133, Jul. 2024, doi: 10.1016/j.engappai.2024.108082.

[18] G. Ke *et al.*, "LightGBM: a highly efficient gradient boosting decision tree," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017, pp. 1–9.
[19] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, vol. 13-17-Augu, pp. 785–794, doi: 10.1145/2939672.2939785.
[20] D. Tamayo *et al.*, "A machine learns to predict the stability of tightly packed planetary systems," *The Astrophysical Journal Letters*, vol. 832, no. 2, Dec. 2016, doi: 10.3847/2041-8205/832/2/L22.
[21] C. Adam-Bourdarios, G. Cowan, C. Germain-Renaud, I. Guyon, B. Kégl, and D. Rousseau, "The Higgs machine learning challenge," *Journal of Physics: Conference Series*, vol. 664, no. 7, Dec. 2015, doi: 10.1088/1742-6596/664/7/072015.
[22] E. O. Dengerud, "Global models for time series forecasting with applications to zero-shot forecasting," Master's thesis, NTNU, 2021.
[23] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas, "Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines," *Ore Geology Reviews*, vol. 71, pp. 804–818, Dec. 2015, doi: 10.1016/j.oregeorev.2015.01.001.
[24] X. Du, H. Xu, and F. Zhu, "Understanding the effect of hyperparameter optimization on machine learning models for structure design problems," *Computer-Aided Design*, vol. 135, Jun. 2021, doi: 10.1016/j.cad.2021.103013.
[25] P. Liashchynskyi and P. Liashchynskyi, "Grid search, random search, genetic algorithm: a big comparison for NAS," *arXiv preprint arXiv:1912.06059*, 2019.
[26] N. Chandrasekhar and S. Peddakrishna, "Enhancing heart disease prediction accuracy through machine learning techniques and optimization," *Processes*, vol. 11, no. 4, Apr. 2023, doi: 10.3390/pr11041210.
[27] O. Guennioui, D. Chiadmi, and M. Amghar, "Global stock price forecasting during a period of market stress using LightGBM," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 19–27, Jan. 2024, doi: 10.12785/ijcds/150103.
[28] Casablanca Stock Exchange, "Data/market data/instruments interface," *Casablanca-bourse.com*, https://www.casablanca-bourse.com/en/instruments (accessed Jan. 17, 2024).
[29] M. C. Jensen, "Optimal utilization of market forecasts and the evaluation of investment performance," *SSRN Electronic Journal*, 2003, doi: 10.2139/ssrn.350426.
[30] S. Yousfi, D. Chiadmi, and M. Rhanoui, "Smart big data framework for insight discovery," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 9777–9792, Nov. 2022, doi: 10.1016/j.jksuci.2021.12.009.

BIOGRAPHIES OF AUTHORS

Omaima Guennioui after completing her degree in computer science and financial mathematics from ENSIMAG (France) in 2014, she gained valuable experience in portfolio management and allocation strategy. Currently, she is pursuing a Ph.D. program at the Mohammadia School of Engineering, Mohammed V University in Rabat. With research interests spanning finance, AI, and computer networks, she aims to merge these fields to create solutions that bridge technology and finance. She can be contacted at email: omaimaguennioui@research.emi.ac.ma.

Dalila Chiadmi is a full professor at Mohammadia School of Engineers -EMI-, Mohammed V University of Rabat, Morocco. She got her PhD degree in computer science from EMI in 1999. Her main research interests focus on the field of data (Open, linked and big data, and big graphs). She is currently leading the SIP team research at EMI and she was appointed leader of the Moroccan Chapter of the Arab Women in Computing Anita Borg Institute Systers' community in 2012. She has co-authored a book and over 100 refereed book chapters, conference and journal publications. She has served in many conferences as chair, member of the program committee, or reviewer. She can be contacted at email: chiadmi@emi.ac.ma.

Mustapha Amghar received his Ph.D. in 1995 in industrial computing from Liege University in Belgium. He is currently a full professor in the Electrical Department at Mohammadia School of Engineering, Mohammed V University in Rabat. For over 30 years, he has been completing counselling missions for public and private institutions. His research interests include control systems engineering, discrete events systems, electronic engineering, and aerospace engineering. He has also confirmed experiences in global navigation satellite systems, Galileo, GPS, survey engineering, localization, and tracking. He can be contacted at email: amghar@emi.ac.ma.