Improving Global Stock Market Prediction with XGBoost and LightGBM Machine Learning Models

Omaima Guennioui*, Dalila Chiadmi, Mustapha Amghar

Mohammadia School of Engineering, Mohammed V University in Rabat, Rabat, Morocco.

Abstract: Predicting stock market prices has been a challenging task for analysts and researchers over the years due to the numerous factors that affect stock prices, making it difficult to achieve accurate predictions. This paper focuses on the application of machine learning techniques, specifically Light Gradient-Boosting Machine (LightGBM) and Extreme Gradient Boosting (XGBoost), to forecast stock market trends. Furthermore, it introduces a hybrid model called Global XGBoost-LightGBM, which combines these individual models. The study adopts a global approach by consistently applying parameter settings across multiple series and evaluates the models using a significant number of stocks. The effectiveness of this approach is validated through experimentation in two distinct stock markets: US and Morocco. Experimental results clearly demonstrate that the integration of LightGBM and XGBoost enhances the ability to identify and interpret fluctuations in stock prices, enabling these models to effectively address the challenges associated with stock market prediction.

Keywords: Global Forecasting Models, Machine learning, Regression, Stock forecasting.

1. INTRODUCTION

Effective investment selection is crucial for organizations like pension funds, as their managed funds have a limited viability horizon. To assist in decision-making, implementing stock market prediction models can aid managers in making better investment choices. By doing so, they can improve their portfolio performance and extend the viability horizon of their managed funds. Predicting stock prices offers the potential benefit of enabling investors and managers to make informed decisions, helping them anticipate future movements in the stock market and identify potential opportunities for profit.

In most cases, two primary methods are employed for predicting the direction of stock prices: fundamental analysis and technical analysis. Fundamental analysis evaluates the intrinsic value of a stock by conducting an in-depth analysis of the company's balance sheet, strategic initiatives, and microeconomic indicators. On the other hand, technical analysis uses technical indicators and charts to identify patterns and signals of future price fluctuations.

Over the past few years, the implementation of machine learning methods increased significantly with a potential use in many domains for various purposes (Sarker, 2021). The financial industry is one of the fields where machine learning has made significant progress and found potential applications in many areas (Hoang & Wiegratz, 2022). Machine learning methods have been widely validated for stock price forecasting (Henrique, Sobreiro & Kimura, 2019) (Strader et al., 2020), but most research has focused on implementing local models with unique parameter sets for each individual stock. Recently, there have been promising results in global forecasting models trained on multiple time series data (Hewamalage, Bergmeir & Bandara, 2021). A global model shares the same set of parameters for all series, while a local model uses distinct parameter sets for each individual series (Januschowski et al., 2020). Compared to local models, global forecasting models have the ability to acquire cross-series information during model training, enabling them to manage complexity and avoid over-fitting on a global scale (Montero-Manso & Hyndman, 2021). The abundance of time series data related to the stock market makes this approach highly interesting for predicting stock market trends. Using global forecasting models that are trained on multiple, related time series data can improve prediction accuracy.

Light Gradient-Boosting Machine (LightGBM) and Extreme Gradient Boosting (XGBoost) are both gradient trees boosting machine learning algorithms that are designed to handle large and complex datasets (Chen & Guestrin, 2016) (Ke et al., 2017). XGBoost uses an ensemble learning approach that builds a sequence of decision trees to make predictions, while LightGBM uses a leaf-wise growth strategy to build decision trees. Both algorithms excel in terms of speed and accuracy, rendering them well-suited for machine learning tasks in areas like finance.

This article investigates the effectiveness of employing LightGBM and XGBoost models in predicting stock prices. The research adopts a global perspective, employing both models to forecast a substantial number of stocks in two distinct and contrasting stock markets: the US and the Moroccan market. These two markets differ significantly in terms of maturity and complexity. Moreover, we introduce a hy-

*Address correspondence to this author at the Mohammadia School of Engineers, Av. Ibn Sina, Rabat (Morocco); Tel: (212) 537 68 71 50;
E-mail: omaimaguennioui@research.emi.ac.ma

brid model named Global XGBoost-LightGBM, aiming to assess its performance in comparison to individual models. Finally, our findings are compared with other models such as Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and Convolutional Neural Networks (CNN). Our main contributions in this study are summarized as follows:

1. We have identified two models based on LightGBM and XGBoost that demonstrate strong performance in global stock market forecasting.
2. We introduce a hybrid model that combines individual models and investigate its responsiveness and predictive capabilities for stock prices.
3. Our research presents an adaptive global approach capable of effectively predicting a diverse range of stocks in various markets.
4. To assess the effectiveness of our proposed models, we evaluate their performance using real data extracted from two distinct stock markets: the US and Moroccan market.
5. Finally, we pay special attention to analysing the results and drawing meaningful insights from our experiments, providing valuable implications for future research.

The rest of this paper is organized as follows. The second section reviews related works on stock price prediction, global forecasting models and gradient tree boosting machine learning algorithms. In section 3, we describe the methodology applied, followed by our experimental results, analysis, and discussion in Section 4. In Section 5, we summarize some conclusions and suggest directions for future research.

2. RELATED WORKS

Numerous studies have explored the application of machine learning techniques for predicting stock market trends. Guo, Li and Xu (2021) developed a hybrid financial time series model called LSTM_LightGBM, which uses data such as opening, closing, highest, lowest prices, trading volume, and adjusted closing price to make predictions. They found their model to be stable and effective in forecasting stock price fluctuations. Similarly, Yun, Yoon and Won (2021) presented a hybrid GA-XGBoost prediction system that incorporates advanced feature engineering processes and optimal feature set selection. Their research demonstrated that incorporating technical indicators through feature expansion enhances prediction performance, and the GA-XGBoost algorithm can produce an optimal feature set with fewer features. Vijh, Chandola, Tikkiwal and Kumar (2020) employed two widely used machine learning methods, Artificial Neural Network (ANN) and Random Forest (RF), to forecast the next day's closing stock prices of five companies from various sectors. They generated new variables from the financial data (Open, High, Low, and Close prices) of each company to serve as inputs to the models. The results revealed that both ANN and RF models effectively predict the next day's closing price for the stocks. Hu, Qin, and Zhang (2023) conducted a study on stock price trend prediction by selecting twelve technical indicators through correlation analysis. They utilized three models: random forest, logistic regression, and LightGBM to forecast the prices of three selected stocks. The models' performance was compared using metrics such as accuracy, recall, precision, and f1 value across different time windows. The findings provided valuable insights into enhancing both long-term and short-term forecasting accuracy. The study highly recommended LightGBM as the preferred model for medium and long-term share price forecasting due to its superior performance. However, many of these studies and existing literature heavily rely on stock-specific data as inputs for their predictive machine learning algorithms. This approach might overlook crucial market-related information and other relevant factors, which could potentially restrict the accuracy and efficiency of the models.

In recent years, global time series models have been developed to capitalize on similarities among related time series. Bojer and Meldgaard (2021) conducted a review of six recent Kaggle forecasting competitions, comparing them to the M4 Time Series Forecasting Competition. The study emphasizes that the forecasting community can gain valuable insights from the Kaggle community, particularly in forecasting daily and weekly business time series. Ensemble models consistently outperformed local single models, reaffirming the findings from the M4 competition. Remarkably, machine learning methods surpassed conventional time series and statistical methods in the recent Kaggle competitions, attributed to their use of external information for cross-learning and modeling the effects of exogenous factors. The top-ranked solutions in Kaggle competitions and the top two solutions in the M4 competition relied on either Gradient Boosting Decision Trees (GBDT) or neural networks. Similarly, Makridakis, Spiliotis, and Assimakopoulos (2022), based on M5 Kaggle's Forecasting Competitions, identified three key findings for forecasting: (1) LightGBM demonstrates remarkable accuracy in predicting retail sales, (2) incorporating external adjustments and explanatory variables improves precision, and (3) utilizing cross-learning and cross-validation is beneficial. Additionally, they highlighted the dominance of machine learning methods in recent competitions, underlining the importance of merging statistics and data science. In this article, we have implemented the suggestions and recommendations provided, specifically those aimed at improving the accuracy of our forecasting.

In conclusion, the review of existing literature highlights the limited utilization of ensemble learning algorithms and global forecasting models in the domain of stock market prediction. In our study, we aim to address this gap by employing a global forecasting approach and utilizing ensemble techniques to develop global forecasting models. Our methodology involves evaluating both individual models and combined model on two distinct stock markets. To validate the effectiveness of our approach, we conducted a comparative analysis with other models known for their accurate time series prediction capabilities.

3. MATERIALS AND METHODOLOGY

3.1. Model Dataset

We retrieved the data from Bloomberg and Casablanca Stock Exchange website (www.casablanca-bourse.com) for the period from January 2012 to January 2023. The main variables of interest were the daily closing prices of individual stocks and their corresponding sectors.

In our research, we concentrated on assessing the effectiveness of our models in two distinct stock markets. The first being the US stock market, representing a mature market characterized by high levels of volatility and returns. The second being the Moroccan stock market, representing an emerging stock market with diverse challenges in the prediction of stock prices.

We chose to study the US stock market through its index, The Dow Jones Industrial Average (DJIA), which is a stock market index that tracks the performance of thirty large companies listed on various stock exchanges in the United States. To do so, we specifically selected stocks from two major sectors within this index: Financials and Information Technology. According to data from January 2023, these two sectors collectively account for more than 35% of the total composition of the DJIA.

Concerning the Moroccan stock market, there are three sectors that collectively account for over 50% of the market's overall movement. Based on data from January 2023, these sectors consist of the banking index, playing a crucial role in the Moroccan economy, the building and construction materials index, reflecting the growth in the country's construction industry, and the telecommunication index. Our study specifically focuses on these three major sectors.

Table **1** and **2** provide the details of the stocks selected for each stock market.

Table 1. Stocks selected from the US market.

Sector	Members
Financials	GOLDMAN SACHS GROUP
	VISA
	TRAVELERS COS
	AMERICAN EXPRESS
	JPMORGAN CHASE
Information Technology	MICROSOFT CORP
	SALESFORCE
	APPLE
	INTERNATIONAL BUSINESS MACHINES
	CISCO SYSTEMS
	INTEL CORP

Table 2. Stocks selected from the Moroccan market.

Sector	Members
Banking	ATTIJARIWAFA BANK
	BANK OF AFRICA
	BCP
	BMCI
	CDM
	CIH
Building and Construction Materials	AFRIC INDUSTRIES SA
	ALUMINIUM DU MAROC
	CIMENTS DU MAROC
	COLORADO
	JET CONTRACTORS
	LAFARGEHOLCIM MAR
	SONASID
	TGCC S. A
Telecommunication	ITISSALAT AL-MAGHRIB

Note that our study excluded companies listed after 2012, such as TGCC S.A (Building and Construction Materials, Moroccan Stock Market) as there was insufficient data available.

3.2. Background: Machine learning algorithms

3.2.1. XGBoost

XGBoost by Chen and Guestrin (2016) is an extensively used open-source software library designed to implement gradient boosting machine algorithms. As one of the most popular boosting tree algorithms in the data science community, it has gained significant popularity due to its high problem-solving performance and minimal feature engineering requirements (Tamayo et al., 2016). XGBoost is an enhanced framework of the GBRT model, which is a boosting model comprising a series of basic regression trees built using a sequential ensemble technique. The model can adaptively add more trees to increase its capacity. Chen and Guestrin introduced a regularization term to the conventional gradient boosted regression Trees loss function in XGBoost to penalize model complexity and prevent overfitting.

XGBoost's large adoption in the industry stems from its ability to handle various structured and unstructured datasets. Furthermore, XGBoost has consistently given high performance in diverse data science tasks, including classification, regression, and ranking.

3.2.2. LightGBM

LightGBM, a machine learning algorithm introduced by Microsoft in 2017, belongs to the gradient boosting decision tree family. It is specifically designed for efficiently training large-scale supervised learning tasks such as regression, classification, and ranking. LightGBM achieves high accuracy in these tasks by iteratively building decision trees and optimizing the leaf-wise split strategy, selecting the optimal feature value for each tree leaf to maximize the overall accuracy gain (Ke et al., 2017).

Table 3. Hyperparameter tuning details for XGBoost and LightGBM.

Algorithm	Hyperparameters	Meanings	Search Ranges
XGBoost	learning_rate	Shrinkage coefficient of each tree	(0.05, 0.1, 0.2)
	max_depth	Maximum depth of a tree	(3, 5, 7)
	n_estimators	Number of iterations	(150, 200, 250)
	subsample	sub-sample size	(0.8, 0.9, 1.0)
	colsample_bytree	Random sampling ratio of features	(0.8, 0.9, 1.0)
	gamma	Minimum loss function degradation required for node splitting	(0, 0.1, 0.2)
	reg_alpha	Regularization on weights	(0, 0.1, 0.2)
	reg_lambda	Regularization on weights	(0, 0.1, 0.2)
LightGBM	n_estimators	Number of iterations	(100, 200, 300)
	learning rate	Shrinkage coefficient of each tree	(0.01, 0.05, 0.1)
	num_leaves	Maximum number of leaves in a tree	(10, 50, 100)
	max_depth	Maximum tree depth	(-1, 5, 10)
	min_child_samples	Minimum number of samples required to create a leaf node	(10, 20, 30)
	subsample	sub-sample size	(0.8, 0.9, 1.0)
	colsample_bytree	Random sampling ratio of features	(0.8, 0.9, 1.0)
	reg_alpha	Regularization on weights	(0, 0.1, 0.5)
	reg_lambda	Regularization on weights	(0, 0.1, 0.5)

A primary advantage of LightGBM is its speed and scalability, as it can handle large datasets with millions of examples and features. It accomplishes this by employing several techniques. Firstly, it uses a histogram-based approach for feature discretization, quickly binning continuous features into discrete values. Secondly, it adopts a leaf-wise data layout instead of a level-wise layout, only computing the gradients and hessians for examples within the same leaf, thus reducing computational cost. Lastly, LightGBM supports parallel and distributed training, allowing it to scale to even larger datasets.

3.3. Hyperparameters Tuning

Numerous researchers have discovered that hyperparameters can significantly influence the final performance of ML models (Rodriguez-Galiano et al., 2015) (Du, Xu and Zhu, 2021). Therefore, determining the optimal hyperparameters is crucial for maximizing the performance of ML models. Both XGBoost and LightGBM models have various internal hyperparameters that can affect their prediction accuracy, but their default values may not be ideal for stock price prediction. To identify the best combination of hyperparameters for our specific dataset, we employed a hyperparameter tuning method called Grid Search. This approach involves systematically assessing the model's performance for each hyperparameter combination within a predefined grid, aiming to find the hyperparameter set that delivers the best model performance (Liashchynskyi & Liashchynskyi, 2019). By evaluating all potential hyperparameter combinations, we could select the optimal set that produced the best model performance. Grid search is a widely used technique for optimizing machine learning models, and in our study, we combined it with cross-validation to improve the generalization performance of both XGBoost and LightGBM models.

The specific details of the hyperparameter tuning can be found in Table **3**, which provides information about the hyperparameters and their respective meanings. Furthermore, we established a search range for each hyperparameter to explore various values during the tuning procedure.

Taking into consideration the complexity of the American stock market and its heightened volatility, and since we believe that factors influencing stock movements in the United States are also relevant to the Moroccan market. We opted to conduct our parameter tuning using data from this market and subsequently apply the tuned parameters for predicting the Moroccan stock market. Furthermore, the parameter selection was made in a way that the tuning process can comprehensively explore various aspects of the two model's behaviours. The chosen values within each parameter range cover a diverse set of options, enabling the tuning process to consider a range of settings, including both conservative and alternative approaches.

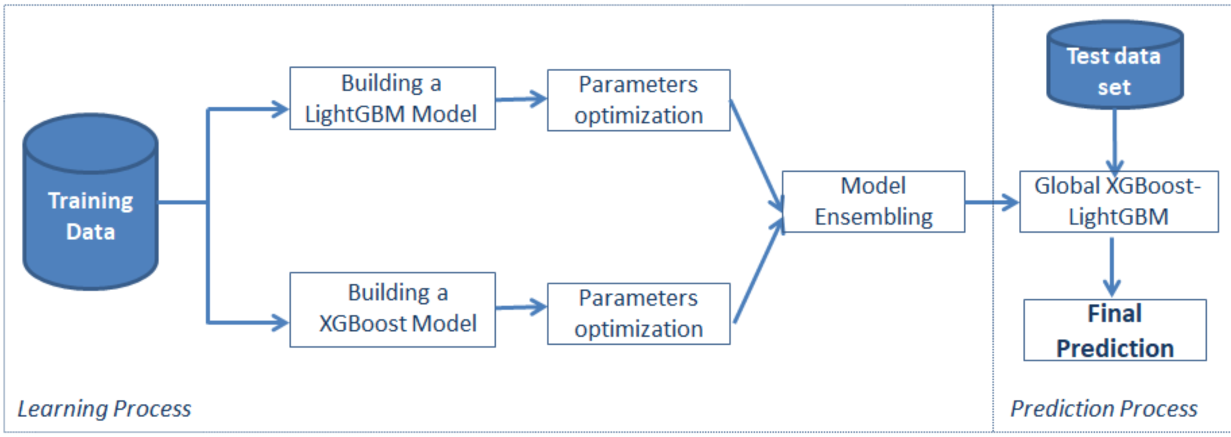

Fig. (1). Applied process chart.

3.4. Global XGBoost-LightGBM

Global XGBoost-LightGBM is a hybrid model that adopts a global approach for predicting stock prices, where the final results are assembled from the predictions of each individual model: XGBoost and LightGBM. Ensemble learning is a machine learning technique that combines the predictions of multiple diverse and independent models to improve overall performance and reduce generalization error.

In our experiment, we employed a voting ensemble method that uses multiple models to enhance system performance for both classification and regression problems. Voting ensembles for regression problems are referred to as voting regressors (VRs),

where estimators from all models are averaged to obtain a final estimate. There are two voting methods: average voting (AV) and weighted voting (WV). In our particular situation, as XGBoost and LightGBM produced comparable outcomes, we chose to adopt the average voting approach (Sagi & Rokach, 2018).

It is important to mention that prior to combining XGBoost and LightGBM, we performed hyperparameter tuning for both models.

4. RESULTS AND DISCUSSION

4.1. Datasets & Baselines

The aim of this study is to forecast the next day's closing price for a selected group of stocks. The dataset used comprises daily closing prices for these stocks throughout the study period, with adjustments made for bank holidays and alignment across stocks. To train the prediction model, the raw dataset was partitioned into training and testing sets at a ratio of 0.8 to 0.2, respectively.

The performance of our proposed model was compared with the performance of deep learning models that are known for their effectiveness in time series forecasting (Guennioui, Chiadmi & Amghar, 2022). Specifically, Long Short-Term Memory (LSTM) networks, a type of Recurrent Neural Networks, were chosen due to their capacity to retain information over extended periods. Convolutional Neural Networks (CNN) were also considered, as they can detect temporal and spatial patterns in high-dimensional data. Additionally, Gated Recurrent Unit (GRU), another type of recurrent neural network, was included in our analysis, given its growing popularity in time series prediction tasks.

For the programming language, we utilized Python, carrying out the development on Google Colab. For the deep learning models, TensorFlow and Keras packages were employed. It's important to note that we did not employ any GPU or cloud computing resources during our experimentation.

4.2. Performance Metrics

We evaluate prediction performance of our regression problem with three different evaluation metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and R-squared (R2). Calculated as follow:

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

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \tilde{y}_j| \tag{2}$$

$$R2 = \sqrt{1 - \frac{\sum_{j=1}^{n}(y_j - \tilde{y}_j)^2}{\sum_{j=1}^{n}(y_j - \bar{y}_j)^2}} \tag{3}$$

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.

4.3. Experimental Results

To prevent issues such as underfitting or overfitting that can arise from manually specifying parameters, it is necessary to tune models' parameters. In our study we use GridSearchCV for parameter tuning. It employs an exhaustive search approach, testing every possible combination of parameters from the given candidates, and selects the best-performing parameter set as the final result. Thus, it ensures the accuracy and reliability of results. The evaluation is based on the search results of different parameter combinations, ultimately leading to an optimal result within the specified parameter range that best fits the model. The results of our parameter tuning for XGBoost and LightGBM are summarized in Table **4**:

Table 4. Hyperparameter Tuning Results.

Algorithm	Hyperparameters	Optimal Values
XGBoost	learning_rate	0.2
	max_depth	3
	n_estimators	200
	subsample	0.9
	colsample_bytree	0.9
	gamma	0
	reg_alpha	0
	reg_lambda	0.1
LightGBM	n_estimators	300
	learning rate	0.1
	num_leaves	10
	max_depth	-1
	min_child_samples	10
	subsample	0.8
	colsample_bytree	0.8
	reg_alpha	0.0
	reg_lambda	0.5

Baseline models were constructed for each stock in the selected group using a local approach, where past observations were used to develop the model. The reported prediction error is the average of the errors calculated for each individual stock prediction.

Table **5** presents the evaluation metrics for the prediction results of different models, including RMSE, MAE, and R2 score. Among the tested models, parameter optimization was employed to achieve the highest level of accuracy.

The table highlights the performances of different models applied to our data in the two selected stock markets: US and Morocco.

For stock predictions in the US market, it is observed that the G-LightGBM, G-XGBoost, and Global XGBoost-LightGBM methods achieve similar performances, with a slight improvement in accuracy when using the model built by combining LightGBM and XGBoost. The RMSE and MAE calculated for the three models, G-LightGBM, G-XGBoost, and Global XGBoost-LightGBM, are very low (around 0.018 for RMSE and 0.0133 for MAE). Furthermore, all three models exhibit a high R2 score, indicating the ability to account for about 99% of the variance in the target variable. These results indicate that these methods are accurate in predicting stocks in the US market. As for the methods based on neural networks (LSTM, GRU, and CNN), lower performances are observed. The RMSE and MAE values for these methods are higher (0.0188 / 0.0144 for LSTM, 0.0255/ 0.0201 for GRU, and 0.0194/0.0146 for CNN). And the R2 score is lower for LSTM, GRU, and CNN (0.967, 0.9913, and 0.963, respectively), indicating that these models explain less variance in the data compared to the proposed methods. Additionally, as these models are designed to build individual models for each dataset, they require significant computational resources.

Regarding stock predictions in the Moroccan market, the G-LightGBM, G-XGBoost, and Global XGBoost-LightGBM methods achieve similar results to those in the US market, with very low RMSE values (0.0192, 0.0195, and 0.0191, respectively) and very low MAE values (0.0119, 0.0124, and 0.0119, respectively). The R2 score is also high for these

Table 5. Performance Results.

Market	Method	Approach	RMSE	MAE	R2 score
US	G-LightGBM	Global	0.0188	0.0140	0.989
	G-XGBoost	Global	0.0187	0.0137	0.989
	Global XGBoost-LightGBM	Global	0.0180	0.0133	0.990
	LSTM	Local	0.0188	0.0144	0.967
	GRU	Local	0.0255	0.0201	0.913
	CNN	Local	0.0194	0.0146	0.963
Morocco	G-LightGBM	Global	0.0192	0.0119	0.991
	G-XGBoost	Global	0.0195	0.0124	0.991
	Global XGBoost-LightGBM	Global	0.0191	0.0119	0.991
	LSTM	Local	0.0194	0.0138	0.978
	GRU	Local	0.0197	0.0143	0.977
	CNN	Local	0.0192	0.0135	0.979

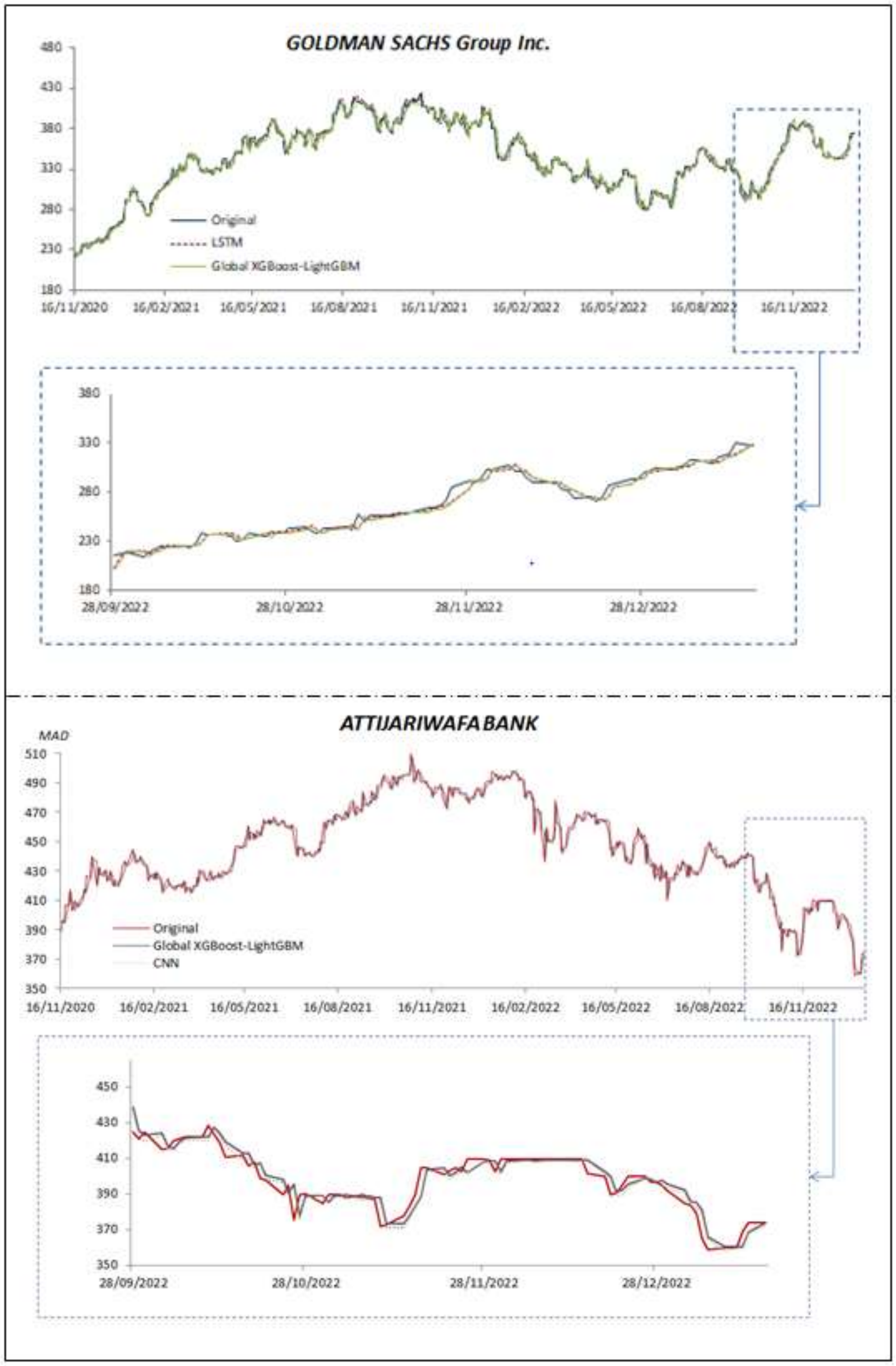

Fig. (2). Predicted closing price of Goldman Sachs and Attijariwafa Bank.

three methods (0.991), indicating an accurate predictive capacity in the Moroccan market. As for the methods based on neural networks, lower performances are observed compared to tree-based methods. Furthermore, The RMSE and MAE values for LSTM, GRU, and CNN are all close, indicating a similar prediction error for these methods.

These results suggest that, for both markets, the proposed models: G-LightGBM, G-XGBoost, and especially their ensemble, Global XGBoost-LightGBM, outperform LSTM, GRU, and CNN models in terms of both the average error of predictions and the proportion of variance explained. Additionally, we observed that, under the evaluated conditions, there is no benefit in utilizing the hybrid model in the Moroccan market compared to the US market. Moreover, the proposed models were designed to predict a significant portion of the markets and have succeeded in doing so. Finally, optimizing the parameters in the US market and transferring these same parameters to the Moroccan market has yielded accurate results. This suggests the possibility of generalizing to stock markets belonging to different economies and influenced by the same factors driving stock movements.

Fig. (**2**) showcases a composite chart of the best-performing algorithms from both the baseline and the proposed models. It displays the prediction results for two major banks in the US and Moroccan markets (Goldman Sachs and Attijariwafa Bank). The predictions are made on the test set. A specific section of the data is also highlighted and presented in the figure for a more detailed view.

5. CONCLUSION

Predicting financial asset prices, such as stocks, is a critical task in economic management. As we enter the era of big data, with an overwhelming amount of information available, there's an increasing trend towards employing computer algorithms to address challenges associated with market prediction. This study investigates the effectiveness of utilizing LightGBM and XGBoost, as well as a hybrid model that combines both, through a global forecasting approach for stock price prediction. To demonstrate the efficacy of our approach, a comparison with deep learning models such as CNN, LSTM and GRU is conducted. The evaluation was performed in two different stock markets.

There are several directions that can be explored following this project. For instance, the approach could be extended to cover other diverse markets. Additionally, as an enhancement to the proposed model, the ensembling method could be further developed to better address the challenges of stock price prediction.

REFERENCES

Adam-Bourdarios, C., Cowan, G., Germain-Renaud, C., Guyon, I., Kégl, B., &

Rousseau, D. (2015). The higgs machine learning challenge. Journal of Physics: Conference Series, 664(7). https://doi.org/10.1088/1742-6596/664/7/072015

Bojer, C. S., & Meldgaard, J. P. (2021). Kaggle forecasting competitions: An

Overlooked learning opportunity. International Journal of Forecasting, 37(2). https://doi.org/10.1016/j.ijforecast.2020.07.007

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System.

Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016, 785–794. https://doi.org/10.1145/2939672.2939785

Guennioui Omaima, Chiadmi Dalila, Amghar Mustapha. (2022). Systematic mapping study on Applications of deep learning in stock market prediction. The 10th International Conference on Innovation and New Trends in Information Technology. In press.

Guo, Y., Li, Y., & Xu, Y. (2021). Study on the application of LSTM-LightGBM Model in stock rise and fall prediction. MATEC Web of Conferences, 336, 05011. https://doi.org/10.1051/MATECCONF/202133605011

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. Expert Systems with Applications, 124, 226–251.

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Global Models for Time Series Forecasting: A Simulation Study. Pattern Recognition (Vol 124).

Hoang, D., & Wiegratz, K. (2022). Machine Learning Methods in Finance: Recent Applications and Prospects. SSRN Electronic Journal.

Hu, Q., Qin, S., & Zhang, S. (2023). 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). https://doi.org/10.2991/978-94-6463-030-5_24

Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M., & Callot, L. (2020). Criteria for classifying forecasting methods. In International Journal of Forecasting (Vol. 36, Issue 1).

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Curran Associates, Inc (Vol. 30).

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. International Journal of Forecasting, 38(4), 1346–1364.

Montero-Manso, P., & Hyndman, R. J. (2021). Principles and Algorithms for Forecasting Groups of Time Series: Locality and Globality. International Journal of Forecasting (Vol. 37, Issue 4).

Liashchynskyi, Petro & Liashchynskyi, Pavlo. (2019). Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. ArXiv abs/1912.06059.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geology Reviews, 71. https://doi.org/10.1016/j.oregeorev.2015.01.001

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. In Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery (Vol. 8, Issue 4).

Sarker, Iqbal. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science. 2. 10.1007/s42979-021-00592-x.

Strader, T. J., Rozycki, J. J., ROOT, T. H., & Huang, Y.-H. J. (2020). Machine Learning Stock Market Prediction Studies: Review and Research Directions. Journal of International Technology and Information Management, 28(4). https://doi.org/10.58729/1941-6679.1435

Tamayo, D., Silburt, A., Valencia, D., Menou, K., Ali-Dib, M., Petrovich, C., Huang, C. X., Rein, H., van Laerhoven, C., Paradise, A., Obertas, A., & Murray, N. (2016). A Machine Learns to Predict the Stability of Tightly Packed Planetary Systems. The Astrophysical Journal, 832(2), L22. https://doi.org/10.3847/2041-8205/832/2/L22

Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock Closing Price Prediction using Machine Learning Techniques. Procedia Computer Science, 167, 599–606. https://doi.org/10.1016/J.PROCS.2020.03.326

Xianping Du and Hongyi Xu and Feng Zhu. (2021). Understanding the Effect of Hyperparameter Optimization on Machine Learning Models for Structure Design Problems. Computer-Aided Design, 135, 103013. https://doi.org/10.1016%2Fj.cad.2021.103013

Yun, K. K., Yoon, S. W., & Won, D. (2021). Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. Expert Systems with Applications, 186. https://doi.org/10.1016/j.eswa.2021.115716

Received: Oct 05, 2023 Revised: Oct 08, 2023 Accepted: Oct 12, 2023

Copyright © 2023– All Rights Reserved
This is an open-access article.