

Forecasting International Stock Market Trends: XGBoost, LSTM, LSTM-XGBoost, And Backtesting XGBoost Models

Hassan Oukhouya ^{1,*}, Hamza Kadiri ², Khalid El Himdi ¹, Raby Guerbaz ²

¹Laboratory LMSA, Department of Mathematics, Faculty of Sciences, Mohammed V University of Rabat, Morocco

²Laboratory MAEGE, Department of SMAEG, FSJES Ain Sebaa, Hassan II University of Casablanca, Morocco

Abstract Forecasting time series is crucial for financial research and decision-making in business. The non-linearity of stock market prices has a profound impact on global economic and financial sectors. This study focuses on modeling and forecasting the daily prices of key stock indices - MASI, CAC 40, DAX, FTSE 250, NASDAQ, and HKEX, representing the Moroccan, French, German, British, US, and Hong Kong markets, respectively. We compare the performance of machine learning models, including Long Short-Term Memory (LSTM), eXtreme Gradient Boosting (XGBoost), and the hybrid LSTM-XGBoost, and utilize the skforecast library for backtesting. Results show that the hybrid LSTM-XGBoost model, optimized using Grid Search (GS), outperforms other models, achieving high accuracy in forecasting daily prices. This contribution offers financial analysts and investors valuable insights, facilitating informed decision-making through precise forecasts of international stock prices.

Keywords Time series, Modeling, Forecasting, Stock market, LSTM, XGBoost, Hybrid model, Backtesting, Grid Search

DOI: 10.19139/soic-2310-5070-1822

1. Introduction

In a laissez-faire world economy, the stock market is a central platform, enabling the exchange of funds between individuals and companies. Renowned for its accessibility and the potential for significant profits, it has long been a magnet for investors searching for lucrative returns and businesses needing capital [1]. Forecasting is crucial in this dynamic landscape, utilising historical data to anticipate forthcoming events. This practice spans diverse disciplines such as environmental science, economics, business, and finance, demonstrating its broad-ranging impact and relevance. The stock market occupies a central position within financial markets, captivating the interest of researchers from both financial and technical backgrounds. The prediction of stock market price trends has emerged as a prominent research area due to its relevance to investors [2]. A significant portion of forecasting challenges revolves around examining time. A time series is a sequential arrangement of observations related to a variable. In this context, the variable under consideration is the stock price. Using sentiment analysis for predicting stock price trends within the context of financial time series helps improve the forecasting of closing prices [3]. In exploring the accuracy of classical, Machine Learning (ML), and Deep Learning (DL) methods in time series forecasting, recent researchers delve into these challenges [4].

The stock price forecasting methods can be divided into two categories: Linear Models such as Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving

*Correspondence to: Hassan Oukhouya (Email: hassan.oukhouya@um5r.ac.ma). Laboratory LMSA, Department of Mathematics, Faculty of Sciences, Mohammed V University of Rabat 4 Avenue Ibn Battouta, B.P. 1014 RP, Rabat, Morocco.

Average (ARIMA) [5], and Nonlinear Models such as Autoregressive conditional heteroskedasticity (ARCH), Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) [6], ML algorithms [7]. Linear models cannot capture the dynamics present in the data, while nonlinear models offer alternative approaches to address this limitation. To address the challenges of accurately predicting stock price direction, we have developed a hybrid model that combines the strengths of both ML and DL methods. The model's framework involves sequential workflow steps, including preprocessing, feature engineering, optimization, prediction, validation, and evaluation. We aim to overcome the complexities involved in stock price prediction and enhance the accuracy of our model. We have thoroughly reviewed the existing research on stock price prediction using ML models, conducting a comprehensive analysis of relevant studies in the field. Ding and Qin [8] introduced an Associated and Deep Recurrent Neural Network (DRNN) model, leveraging the Long Short-Term Memory (LSTM) network model with multiple inputs and outputs. The experimental results demonstrate the superior accuracy of the associated model compared to the other two models in simultaneously predicting multiple values, achieving a prediction accuracy of over 95%. Shen and Shafiq [9] studied Chinese stock market data, employing comprehensive preprocessing and feature engineering techniques. They developed a customized LSTM-based system that outperformed other approaches in accurately predicting stock market price trends. The study emphasized the importance of effective feature engineering for achieving high accuracy in stock market trend prediction. Vuong et al. [10], eXtreme Gradient Boosting (XGBoost) was used for feature selection in high-dimensional time-series data, followed by LSTM for stock price forecasting. Results showed that this approach outperformed the baseline ARIMA model regarding Mean Squared Error (MSE), Root-Mean-Square Error (RMSE), and Mean Absolute Error (MAE) on a Forex dataset from 2008-2018. In a comparative conducted by Oukhouya and El Himdi [11], various ML methods, including XGBoost, Support Vector Regression (SVR), Multilayer Perceptron (MLP), and LSTM, were evaluated for forecasting a daily closing price of the Morocco Stock Index 20 (MSI 20). The results demonstrated that the SVR and MLP models, optimized with Grid Search (GS), have outperformed other models and achieved outstanding accuracy in price prediction.

Some investigators have explored the use of hybrid models to surpass the limitations of single models, with promising results. Various forecasting techniques were explored in a study conducted by the authors [12]. The initial approach employed linear modeling techniques, including classical ARIMA-GARCH and exponential smoothing models. The subsequent approach employed non-linear modeling techniques, specifically Artificial Neural Networks (ANNs), focusing on the MLP model. The results suggest that the MLP model with exogenous variables outperforms the alternative models, highlighting its effectiveness in renewable energy and time series forecasting. The study presented in [13] proposed a hybrid bi-directional LSTM-XGBoost model for accurate energy community load forecasting. By leveraging separate forecasts for general load patterns and peak loads and comprehensively integrating them, the hybrid model demonstrated superior performance compared to conventional methods relying on standard load profiles and LSTM-based forecasts. A separate study [14] introduces a hybrid ML model that combines XGBoost and LSTM to predict advanced truck parking occupancy at 30, 60, 90, and 120-minute intervals. The model demonstrates high accuracy in its predictions, utilizing historical occupancy data and ensuring compatibility with existing truck occupancy detection systems for future forecasting. This paper presents the LSTM-XGBoost, a novel hybrid model that combines LSTM residuals within the XGBoost framework. The proposed model is compared to individual XGBoost and LSTM models to improve the accuracy of predicting fluctuations in six international stock prices. Additionally, by utilizing the skforecast library [15], we enhance the interval prediction of Bootstrap based on XGBoost through backtesting. Our findings demonstrate that the proposed model effectively tracks both stocks' upward and downward movements with its optimized GS approach and enhanced capture ability. This leads to improved prediction performance compared to using a single XGBoost or LSTM model for forecasting multivariate stock prices.

The rest of this article is organized as follows. Section 2 presents the experimental data set of six international stock markets and the data preprocessing techniques employed. The design methodology, including the algorithms utilized and the implementation details of the models, is also described in this section. Section 3 of this article thoroughly analyzes and evaluates the proposed model, including a comparison with alternative models. The findings and their implications are discussed in detail. Finally, Section 4 concludes the paper by summarizing the main findings and suggesting potential avenues for future research.

2. Method

This section describes the methodology for modeling and forecasting six international stock market trends. It includes time series pre-processing for multivariate data, the implementation of XGBoost, LSTM, and hybrid LSTM-XGBoost models, and the architectural design of the research process.

2.1. Handling Multivariate Time Series Data

The historical data used in the experiments consists of multiple time series of daily prices (Open, Low, High, Close prices, and Volume) for six international stock markets across different continents, such as DAX, MASI, HKEX, CAC 40, NASDAQ, and FTSE 250. The series covers the period from March 1st, 2018, to March 1st, 2023, with 1,259 trading days for each stock market series. Several reasons justify the choice of 5 years and including six international stock markets: (1) investors commonly analyze stock market trends using recent data for more relevant insights, (2) assessing the models' performance during and before crisis periods, such as the COVID-19 pandemic, and (3) evaluating our model's performance across different continents and contexts.

The data was collected from [investing.com](https://www.investing.com). Let $X^{\text{stock}}(t) = \{x_t^{\text{stock}_{i,j}}; t \in T\}$ represent a multivariate time series for each stock, where i represents the DAX, MASI, HKEX, CAC 40, NASDAQ, FTSE 250 respectively, and j represents the Open, Low, and High prices, Log_volume respectively. We used the closing price of each share as a forecasting target with a single observation sequence,

$$Y^{\text{stock}_i} = \hat{y}_{t+1}^{\text{stock}_i} \quad (1)$$

The preprocessing step is crucial in time series analysis, where all multiple input series is transformed into a homogeneous numerical array to be compatible with the XGBoost and LSTM models. Data normalization is applied using the min-max formula (Eq. 2) to ensure that features with different scales are processed together.

$$\tilde{x}_t = \frac{x_t^{\text{stock}_{i,j}} - \min(x_t^{\text{stock}_{i,j}})}{\max(x_t^{\text{stock}_{i,j}}) - \min(x_t^{\text{stock}_{i,j}})} \quad (2)$$

This normalization technique facilitates faster convergence of the gradient descent algorithm and maintains the integrity of relationships within the data without introducing bias [16], [17].

2.2. XGBoost Model

The XGBoost model, developed by Li and Zhang [18] and Chen and Guestrin [19], is a robust ML algorithm known for its practicality and speed of execution. It employs gradient-boosted decision trees for efficient feature selection and has shown promising results in stock market forecasting due to its parallelization and scalability capabilities. The equation to predict the target series for sample i can be expressed as:

$$\hat{y}_{t+j} = \sum_{i=1}^I f_i(x_j), \quad f_i \in \mathbf{F}, \quad (j = 1, 2, \dots, N) \quad (3)$$

where \mathbf{F} represents the space of regression trees. The idea of XGBoost is to minimise the generalised objective function (or loss function) can be written as:

$$L_k = \sum_{j=1}^N (y_{t+j} - \hat{y}_{t+j})^2 + \sum_{k=1}^K \left(\gamma T + \frac{1}{2} \alpha \sum_{m=1}^T \omega_m^2 \right) \quad (4)$$

where T consists of the number of leaf nodes, N is the set of all samples in leaf m . The score of leaf m is measured by ω_m . α and γ are parameters of the tree.

2.3. LSTM Model

LSTMs, a subtype of Recurrent Neural Networks (RNNs), effectively address the challenge of exploding and vanishing gradients [20], [21] by incorporating a memory cell and gate mechanism [22]. The LSTM model consists of blocks with input (i_t), forget (f_t), and output gates (o_t). These blocks form a recurrent structure where the output of one block is fed into the next [14]. Using LSTM models in stock market forecasting enables the effective capture of long-term dependencies, resulting in precise predictions. The LSTM cell operates based on four equations, which govern its underlying principle [11]:

$$\begin{aligned} f_t &= \delta(\mathcal{W}_{ef} \cdot (\tilde{x}_t, h_{t-1}) + \mathcal{W}_{cf}C_{t-1} + b_f), \\ i_t &= \delta(\mathcal{W}_{ki} \cdot (\tilde{x}_t, h_{t-1}, C_{t-1}) + b_i), & k = x, h, c \\ C_t &= f_t \cdot C_{t-1} + i_t \cdot \tanh(\mathcal{W}_{lc}(\tilde{x}_t, h_{t-1}) + b_c), & l = x, h \\ o_t &= \delta(\mathcal{W}_{ji} \cdot (\tilde{x}_t, h_{t-1}, C_{t-1}) + b_o), & j = x, h, c \end{aligned}$$

The notation used in the equation includes the weights (\mathcal{W}_{e*} , \mathcal{W}_{k*} , \mathcal{W}_{c*} , \mathcal{W}_{j*} , \mathcal{W}_{l*}) and biases (b_f , b_c , b_i , b_o) for each layer. The hidden layer output (h_t) is calculated as the element-wise multiplication of the output gate (o_t) and the hyperbolic tangent of the cell state (C_t).

2.4. Hybrid LSTM-XGBoost Model

After forecasting the closing prices using the LSTM model, we calculated the residuals by taking the difference between the actual and predicted values at each time step for each stock, as expressed in Eq. 5. To develop our hybrid LSTM-XGBoost model, we utilized the GS algorithm for both training and evaluation processes. The GS algorithm allowed us to systematically explore different combinations of hyperparameters to find the optimal configuration for our hybrid model. This approach enhances the accuracy and performance of our model in predicting stock market trends.

$$\text{Residual}_{t+1}^{\text{stock}_i} = y_{t+1} - \hat{y}_{t+1} \quad (5)$$

The optimal parameters of the hybrid LSTM-XGBoost model are determined using the GS optimization algorithm. The search space for these parameters is detailed in subsection 3.1 of the article.

2.5. Proposed method

In this study, the forecasting of the six stocks is divided into three main sections: 1) Data Collection and Preprocessing, 2) Model Learning and Testing, and 3) Forecasting and Backtesting with Bootstrap. The flowchart in Figure 1 provides an overview of the proposed approach, highlighting the critical functions of each section. These sections collectively form the framework for our research, enabling us to effectively collect, preprocess, train, test, and forecast stock market data. Our research used backtesting with prediction interval bootstrap to forecast multiple series for each stock, including Open, High, and Low prices and log_volume. Using a recursive multi-step approach, we relied on the previous predictions to forecast the Close price series. The skforecast library's Forecaster object facilitated model training and future prediction generation. By conducting backtesting, we assessed the accuracy and reliability of the predictive XGBoost model by applying it retrospectively to historical data, ensuring its effectiveness in forecasting future time series values.

3. Result and Discussion

This section discusses the results of forecasting six international stock markets using three different models. We conduct a thorough comparative analysis to evaluate their performance and assess the XGBoost model's backtesting capabilities. Additionally, we offer detailed insights into the performance measures achieved by each model. All experiments were conducted on a macOS system featuring a 2.4 GHz Intel Core i5 processor and 16 GB RAM. Furthermore, model evaluation and training took place on a Google Colab Pro+ virtual machine utilising eight

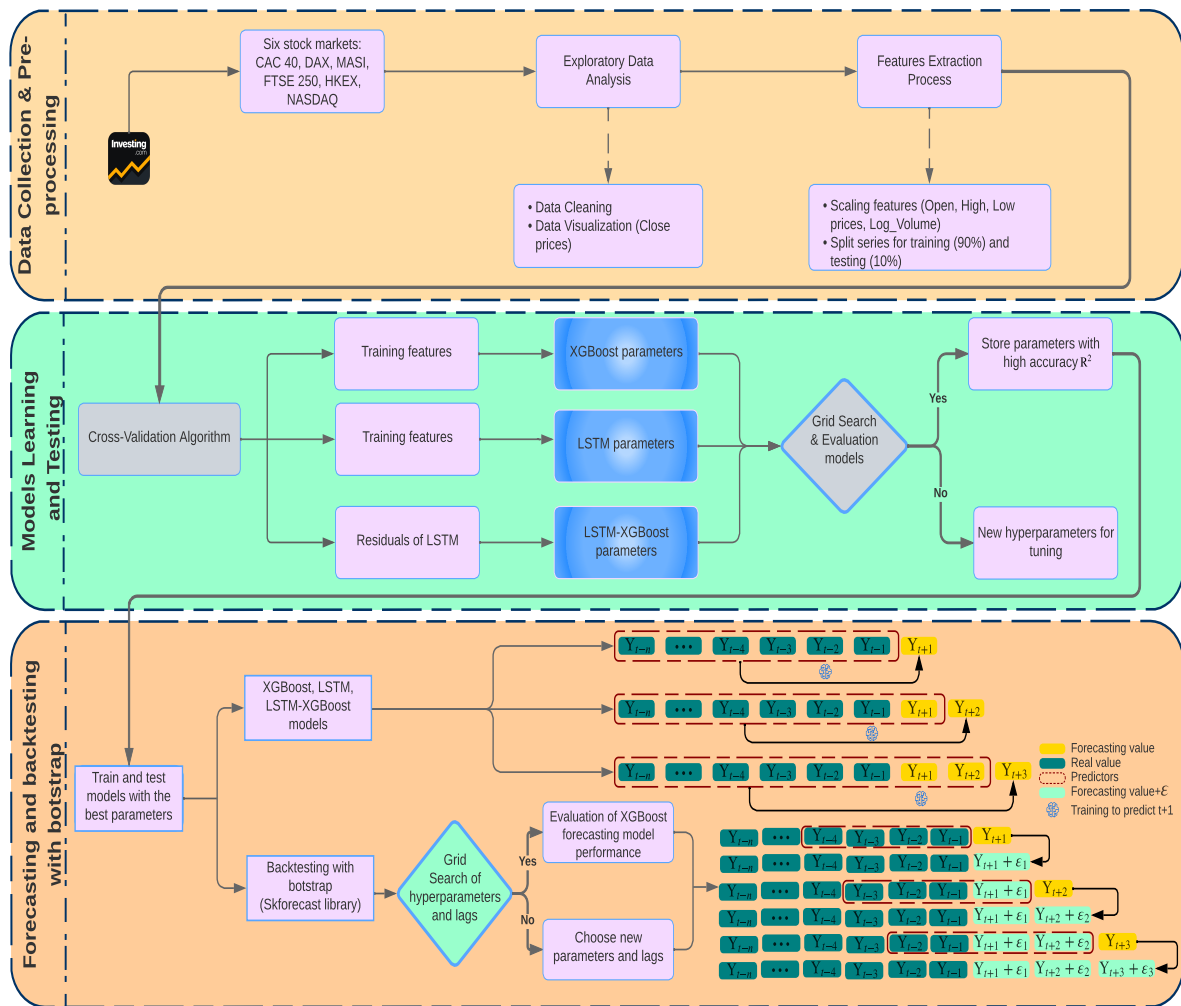


Figure 1. Workflow of this study

TPUs and High-RAM. The LSTM model was implemented using Keras with TensorFlow [23] and XGBoost (v. 1.7.5) with the Scikit-Learn library as the underlying framework [24] within the Anaconda3 environment [25] using Python.

3.1. Models Grid Search Results

The GS procedure utilized a 5-fold blocked time series split cross-validation technique on the training set to discover the optimal hyperparameters for each model. The performance of the models was assessed using MSE as the evaluation metric for multi-step output. The GS results provided the optimal hyperparameters for the XGBoost, LSTM, LSTM-XGBoost, and Backtesting-XGBoost models, summarized in Table 1. Subsequently, the best parameters were applied to train the respective models, and the close pricing of the six stocks was predicted for both the testing and training series.

3.2. Comparative Analysis of Models Performance

The predictive performance of the XGBoost, LSTM, and proposed LSTM-XGBoost hybrid models in forecasting the directional movements of six stock prices (DAX, MASI, HKEX, CAC 40, NASDAQ, and FTSE 250) was

Table 1. The Grid Search Hyperparameter Results for the XGBoost, LSTM, LSTM-XGBoost, and Backtesting-XGBoost Models

Models	Parameters	CAC 40	DAX	MASI	FTSE 250	HKEX	NASDAQ
XGBoost	# of regression trees	1000	100	300	500	500	800
	Maximum regression tree depth	60	30	20	30	10	10
	Boosting the rate of learning	0.03	0.1	0.02	0.02	0.03	0.05
	Minimum reduction of loss	0.00001	0.001	0.01	0.1	0.1	0.001
	Regularization term L1 on weights	0.5	0.3	0.1	0.1	0.09	0.3
	Objective learning function	Squarederror	Squarederror	Squarederror	Squarederror	Squarederror	Squarederror
LSTM	Batch size	32	16	4	8	8	16
	Neurons in the 1st hidden layer	250	900	300	400	150	450
	Neurons in the 2nd hidden layer	400	900	300	500	300	200
	Number of epochs	100	100	100	100	100	100
	Patience	3	7	4	5	4	4
LSTM-XGBoost	# of regression trees	1000	300	300	1000	600	600
	Maximum regression tree depth	2	4	2	5	2	3
	Boosting the rate of learning	0.1	0.03	0.01	0.008	0.02	0.04
	Minimum reduction of loss	0	0	0	0	0.001	0
	Objective learning function	Squarederror	Squarederror	Squarederror	Squarederror	Squarederror	Squarederror
Backtesting	# of regression trees	900	300	700	1000	500	300
	Maximum regression tree depth	1	2	3	3	1	2
	Boosting the rate of learning	0.03	0.04	0.09	0.09	0.03	0.03

evaluated using several performance metrics (see Table 2), including MAE, Mean Absolute Percentage Error (MAPE), RMSE, Theil’s U Statistics, and R^2 . The results of these metrics are presented in Table 3. Let the error of one-step-ahead forecast is defined as $e_t = y_{t+1} - \hat{y}_{t+1|t}$ where \hat{y}_{t+1} denote the forecasted value of the model at time t , y_{t+1} represents the actual value of the test set at time t , \bar{y}_t is the mean value, and n is the length of the time series dataset (including both training and test sets). Each model estimates one of the interval limits to construct a prediction interval using the Backtesting XGBoost model. In our study, the models obtained for $Q = 0.1$ and $Q = 0.9$ create an 80% prediction interval ($90\% - 10\% = 80\%$). The findings presented in Table 3 equivocally establish the superior performance of the hybrid LSTM-XGBoost model across all error measures compared to the benchmark model. These results affirm that the LSTM-XGBoost model exhibits exceptional accuracy when applied to the six stock market indices under scrutiny. Notably, this model delivers precise predictions while effectively managing the trade-off between the training and testing datasets. Additionally, the XGBoost model emerges as a viable and commendable alternative for investors.

Table 2. Formulas for model performance measures.

Metrics	MAE	Theil’s U Statistics	MAPE (%)	RMSE	R^2 (%)
Formulas	$\frac{1}{n} \sum_{t=1}^n e_t $	$\sqrt{\frac{\sum_{t=1}^{n-1} (e_t/y_t)^2}{\sum_{t=1}^n ((y_{t+1}-y_t)/y_t)^2}}$	$\frac{100}{n} \sum_{t=1}^n \left \frac{e_t}{y_t} \right $	$\sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2}$	$1 - \frac{\sum_{t=1}^n e_t^2}{\sum_{t=1}^n (y_t - \bar{y}_t)^2}$

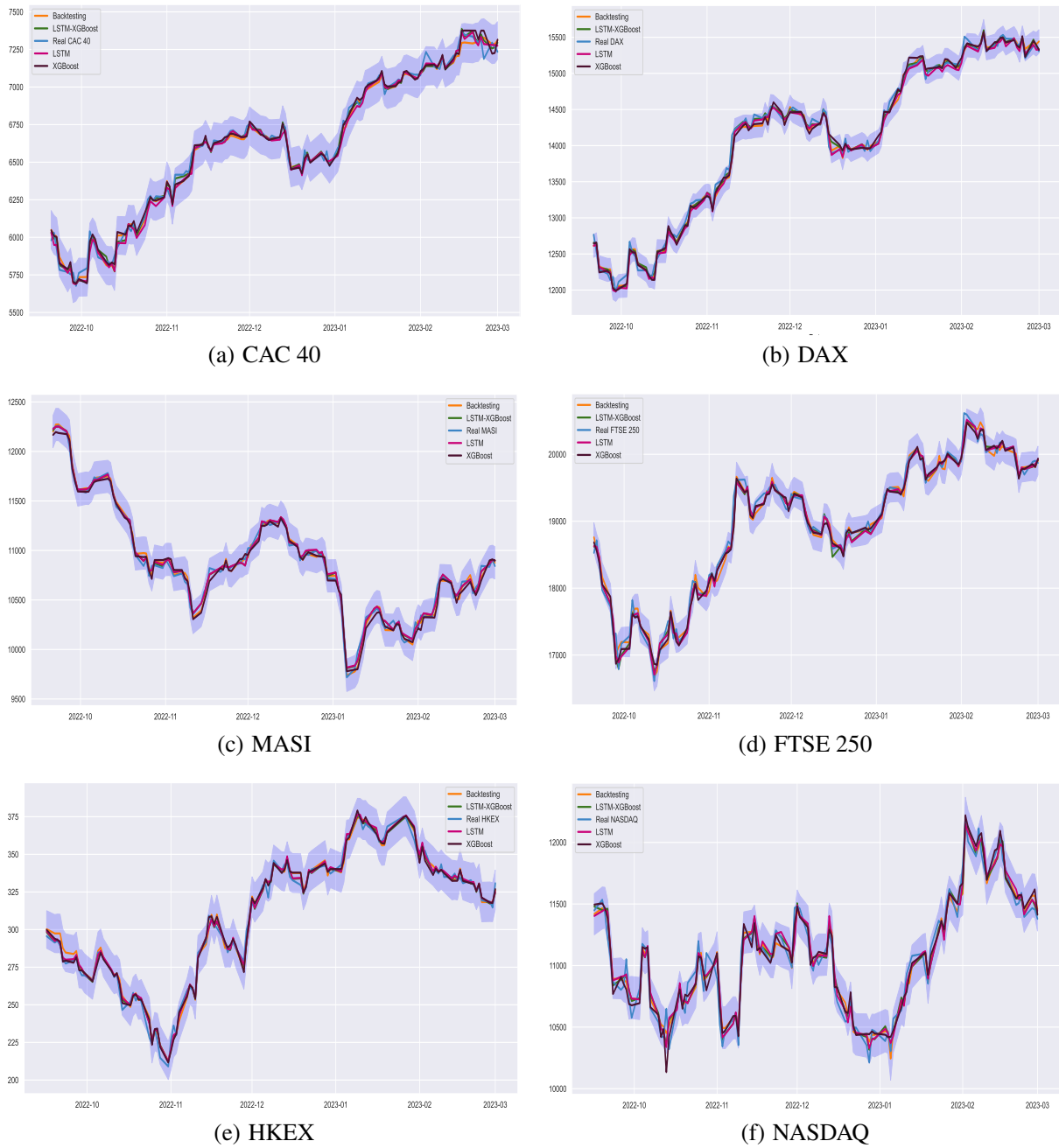


Figure 2. Comparative of Stock Price Forecasting for MASI, CAC 40, DAX, FTSE 250, NASDAQ, and HKEX: XGBoost, LSTM, LSTM-XGBoost, and Backtesting Models Results.

Theil’s U Statistics values close to zero for all models further confirm the excellent forecasting ability of the hybrid model in predicting the trends of the six stock markets compared to the benchmark. Figure 2 provides a visual representation of the performance of the hybrid model and backtesting, highlighting the accuracy and quality of the predictions; Our results reveal that model effectiveness varies by index, with XGBoost excelling

Table 3. Performance comparison of forecasting six stock markets and benchmark models in out-of-sample periods

Models	MAE		MAPE (%)		RMSE		Theil's U Statistics*		R ²	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
XGBoost										
DAX	1.724	62.153	0.013	0.449	3.043	81.732	0.002	0.041	0.999	0.994
MASI	30.76	49.853	0.261	0.464	37.947	64.703	0.027	0.055	0.998	0.984
HKEX	0.381	2.871	0.124	0.970	0.504	3.730	0.420	3.857	0.999	0.992
CAC 40	0.069	31.341	0.001	0.479	0.103	42.393	0.0003	0.096	0.999	0.991
NASDAQ	1.080	75.528	0.011	0.686	1.611	102.150	0.001	0.083	0.999	0.951
FTSE 250	3.578	82.499	0.018	0.443	6.036	118.972	0.001	0.033	0.999	0.986
LSTM										
DAX	54.176	59.731	0.420	0.433	73.096	80.322	0.042	0.040	0.997	0.994
MASI	26.002	32.709	0.225	0.306	34.602	40.162	0.025	0.034	0.998	0.994
HKEX	2.1180	2.787	0.616	0.939	2.898	3.503	2.407	3.617	0.999	0.993
CAC 40	24.561	30.386	0.439	0.469	34.528	40.852	0.104	0.093	0.997	0.992
NASDAQ	60.264	83.205	0.570	0.754	85.667	101.700	0.072	0.083	0.999	0.952
FTSE 250	67.561	74.758	0.348	0.310	91.031	101.420	0.022	0.028	0.998	0.989
LSTM-XGBoost										
DAX	34.952	49.958	0.269	0.365	45.676	67.062	0.026	0.034	0.999	0.996
MASI	24.651	31.706	0.213	0.296	32.584	39.120	0.023	0.033	0.999	0.995
HKEX	1.7360	2.569	0.523	0.867	2.288	3.347	1.904	3.463	0.999	0.994
CAC 40	10.513	25.513	0.186	0.393	13.427	34.349	0.040	0.078	0.999	0.994
NASDAQ	32.437	77.660	0.316	0.705	43.465	95.260	0.037	0.078	0.999	0.957
FTSE 250	45.795	73.350	0.230	0.392	57.837	99.300	0.014	0.028	0.999	0.990

* $\times 10^{-5}$

with the DAX index and LSTM performing optimally with NASDAQ. Overall, the LSTM-XGBoost hybrid model consistently demonstrates high accuracy, showing promise in stock price prediction across diverse indices.

4. Conclusion

This study aimed to assess the performance of different algorithms in predicting stock markets across various continents. Three models, namely XGBoost, LSTM, and LSTM-XGBoost, were employed to evaluate the effectiveness of backtesting performance, specifically in the prediction interval. The experimental findings demonstrated that the proposed hybrid model significantly enhanced the accuracy of forecasting the trends in six international stock markets, surpassing the performance of the individual XGBoost and LSTM models. This research provides valuable insights into the comprehensive exploration of the impact of Backtesting with refit and fixed training size (rolling origin) as well as GS hyper-parameter optimization on the performance of the models. The hybrid model, while showing promise, manifests limitations associated with market volatility, hyperparameter sensitivity, time lag, and offset. These elements necessitate consideration in real-world applications, emphasizing the need for continuous research to enhance its robustness. In future research, we aim to enhance the stock market forecasting model by adding more variables and exploring the potential of other hybrid models, such as CNN-LSTM. We also plan to improve the performance of Backtesting by incorporating exogenous variables. These advancements will improve accuracy and reliability in predicting stock market trends.

Acknowledgement

We thank Professors Tarek Zari, Hicham El Bouanani, and Sophia Faris for their invaluable input and thorough review of this research. We also appreciate the reviewers' contributions in enhancing our work's quality.

REFERENCES

- [1] 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. 115 716, 2021.
- [2] E. Guresen, G. Kayakutlu, and T. U. Daim, "Using artificial neural network models in stock market index prediction," *Expert systems with Applications*, vol. 38, no. 8, pp. 10 389–10 397, 2011.
- [3] M.-L. Thormann, J. Farchmin, C. Weisser, R.-M. Kruse, B. Säfken, and A. Silbersdorff, "Stock price predictions with lstm neural networks and twitter sentiment," *Statistics, Optimization & Information Computing*, vol. 9, no. 2, pp. 268–287, 2021.
- [4] Y. Joulil *et al.*, "Comparing the accuracy of classical and machine learning methods in time series forecasting: A case study of usa inflation," *Statistics, Optimization & Information Computing*, vol. 11, no. 4, pp. 1041–1050, 2023.
- [5] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, "Stock price prediction using the arima model," in *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, 2014, pp. 106–112.
- [6] P. C. S. Bezerra and P. H. M. Albuquerque, "Volatility forecasting via svr-garch with mixture of gaussian kernels," *Computational Management Science*, vol. 14, pp. 179–196, 2017.
- [7] D. Shah, H. Isah, and F. Zulkernine, "Stock market analysis: A review and taxonomy of prediction techniques," *International Journal of Financial Studies*, vol. 7, no. 2, 2019.
- [8] G. Ding and L. Qin, "Study on the prediction of stock price based on the associated network model of lstm," *International Journal of Machine Learning and Cybernetics*, vol. 11, Jun. 2020.
- [9] J. Shen and M. O. Shafiq, "Short-term stock market price trend prediction using a comprehensive deep learning system," *Journal of big Data*, vol. 7, no. 1, pp. 1–33, 2020.
- [10] P. H. Vuong, T. T. Dat, T. K. Mai, P. H. Uyen, *et al.*, "Stock-price forecasting based on xgboost and lstm," *Computer Systems Science & Engineering*, vol. 40, no. 1, 2022.
- [11] H. Oukhouya and K. El Himdi, "Comparing machine learning methods—svr, xgboost, lstm, and mlp—for forecasting the moroccan stock market," in *Computer Sciences & Mathematics Forum*, MDPI, vol. 7, 2023, p. 39.
- [12] H. Ettayyebi and K. El Himdi, "Artificial neural network for forecasting one day ahead of global solar irradiance," *Smart Application and Data Analysis for Smart Cities (SADASC'18)*, 2018.
- [13] L. Semmelmann, S. Henni, and C. Weinhardt, "Load forecasting for energy communities: A novel lstm-xgboost hybrid model based on smart meter data," *Energy Informatics*, vol. 5, no. 1, pp. 1–21, 2022.
- [14] S. Gutmann, C. Maget, M. Spangler, and K. Bogenberger, "Truck parking occupancy prediction: Xgboost-lstm model fusion," *Frontiers in Future Transportation*, vol. 2, p. 693 708, 2021.
- [15] J. Amat Rodrigo and J. Escobar Ortiz, *skforecast*, version 0.9.0, Jul. 2023.
- [16] H. Chung and K.-s. Shin, "Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction," *Neural Computing and Applications*, vol. 32, pp. 7897–7914, 2020.
- [17] P. Devan and N. Khare, "An efficient xgboost-dnn-based classification model for network intrusion detection system," *Neural Computing and Applications*, vol. 32, pp. 12 499–12 514, 2020.
- [18] S. Li and X. Zhang, "Research on orthopedic auxiliary classification and prediction model based on xgboost algorithm," *Neural Computing and Applications*, vol. 32, pp. 1971–1979, 2020.

- [19] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [20] R. Pascanu, T. Mikolov, and Y. Bengio, “On the difficulty of training recurrent neural networks,” in *International conference on machine learning*, Pmlr, 2013, pp. 1310–1318.
- [21] J. Schmidhuber, S. Hochreiter, *et al.*, “Long short-term memory,” *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [22] Y. Wang and X. S. Ni, “A xgboost risk model via feature selection and bayesian hyper-parameter optimization,” *arXiv preprint arXiv:1901.08433*, 2019.
- [23] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow.* ” O’Reilly Media, Inc.”, 2022.
- [24] M. Abadi, A. Agarwal, P. Barham, *et al.*, “Tensorflow: Large-scale machine learning on heterogeneous distributed systems,” *arXiv preprint arXiv:1603.04467*, 2016.
- [25] *Anaconda software distribution*, version Vers. 2-2.4.1, 2020. [Online]. Available: <https://docs.anaconda.com/>.
- [26] S. Selvin, R. Vinayakumar, E. Gopalakrishnan, V. K. Menon, and K. Soman, “Stock price prediction using lstm, rnn and cnn-sliding window model,” in *2017 international conference on advances in computing, communications and informatics (icacci)*, IEEE, 2017, pp. 1643–1647.
- [27] A. Dezhkam and M. T. Manzuri, “Forecasting stock market for an efficient portfolio by combining xgboost and hilbert–huang transform,” *Engineering Applications of Artificial Intelligence*, vol. 118, p. 105 626, 2023.
- [28] S. Deng, X. Huang, Y. Zhu, Z. Su, Z. Fu, and T. Shimada, “Stock index direction forecasting using an explainable extreme gradient boosting and investor sentiments,” *The North American Journal of Economics and Finance*, vol. 64, p. 101 848, 2023.