PREDICTING STOCK PRICE DIRECTION OF EUROZONE BANKS: CAN DEEP LEARNING TECHNIQUES OUTPERFORM TRADITIONAL MODELS?

Bogdan Ionuț ANGHEL, PhD Candidate*

Abstract

Due to market volatility and complex regulations, forecasting stock price movements within the European banking sector is highly challenging. This study compares the predictive performance of Bidirectional Long Short-Term Memory (BiLSTM) and Long Short-Term Memory (LSTM) with traditional models - Extreme Gradient Boosting (XGBoost) and Logistic Regression - in predicting the daily stock price direction of the ten largest Eurozone banks by market capitalization. Utilizing a dataset from January 1, 2000, to May 31, 2024, comprising eight financial and macroeconomic indicators, a comparative analysis of these models was conducted. The findings suggest that traditional machine learning models are more effective than advanced deep learning models for predicting stock price direction in the Eurozone banking sector. The underperformance of LSTM and BiLSTM may be attributed to dataset limitations relative to deep learning requirements.

Keywords: Financial Market, European Banking Sector, Time Series, Prediction

JEL Classification: C53, G17, G12, G21

1. Introduction

Predicting stock price movements within the European banking sector is a complex and challenging task, influenced by market volatility and the intricate regulatory landscape of the Eurozone. Recent advancements in machine learning and deep learning have introduced sophisticated models aimed at capturing the nonlinear and temporal

* *Faculty of International Business and Economics, Bucharest University of Economic Studies, Bucharest, Romania.*

dynamics inherent in the financial markets. This study seeks to evaluate whether Bidirectional Long Short-Term Memory (BiLSTM) and Long Short-Term Memory (LSTM) networks can outperform traditional models: Extreme Gradient Boosting (XGBoost), and Logistic Regression in predicting the direction of stock prices for the 10 largest Eurozone banks by market capitalization.

Financial forecasting necessitates robust models capable of handling complex data patterns and temporal dependencies. Traditional machine learning models, such as Logistic Regression and ensemble methods like XGBoost, have been extensively used due to their interpretability and effectiveness with structured data in classification problems. Deep learning models, particularly recurrent neural networks like LSTM and BiLSTM, offer the potential to model sequential data and capture long-term dependencies, which are essential in time-series forecasting.

The literature reflects a growing interest in applying these advanced models to financial prediction tasks. Despite these advancements, there is a gap in the literature regarding the comparative performance of BiLSTM networks against traditional models like Logistic Regression and XGBoost in the context of the Eurozone banking sector. Moreover, the specific dynamics of European banks, influenced by regional economic policies and market conditions, necessitate a tailored approach to forecasting that accounts for sector-specific characteristics and macroeconomic indicators.

The main contributions of this paper are threefold. First, the study conducts a comprehensive evaluation of BiLSTM, LSTM, XGBoost, and Logistic Regression models in predicting the daily stock price direction of the ten largest Eurozone banks included in the STOXX600 index. Second, the paper enhances the predictive models by incorporating eight carefully selected financial and macroeconomic indicators, following extensive data preprocessing steps. Third, the study assesses the models using key performance metrics, including Area Under the Curve (AUC) and accuracy, to provide a nuanced understanding of each model's strengths and limitations in this specific financial context.

The paper is organized as follows: Section 2 - provides a detailed review of the relevant literature, highlighting previous studies on financial forecasting using machine learning and deep learning models. Section 3 outlines the methodology, including data collection, preprocessing, and the implementation of the predictive models. Section 4 presents the results of the analysis, comparing the performance of each model and discussing the implications. Finally, Section 5 concludes the paper by summarizing the key findings and suggesting directions for future research.

2. Literature review

Forecasting stock movements within the European banking sector is a challenging endeavour, driven by both the high volatility of financial markets and the distinct regulatory dynamics within the Eurozone. The recent literature highlights the application of advanced machine learning methods, deep learning techniques, and hybrid models that aim to capture the nuanced economic and structural dependencies within this sector.

Traditional machine learning models have been widely employed for stock movement prediction. Qiu and Song (2016) demonstrated the efficacy of optimized ANN models, particularly when combined with genetic algorithms, in improving prediction accuracy for stock indices. Zhong and Enke (2019) extended this approach by deploying hybrid machine learning techniques to classify the daily return direction of the S&P 500, using Deep Neural Networks enhanced with Feature Engineering Techniques. More advanced recurrent models, specifically LSTM and BiLSTM, have also proven effective for time-series data with temporal dependencies. BiLSTM models, which capture dependencies from both forward and backward sequences, have shown success in various domains, including stock prediction and agriculture classification problems. For instance, Kwak et al. (2020) explored BiLSTM for classification, indicating the model's capability to utilize multi temporal dependencies effectively. This bidirectional aspect provides a broader context that can be particularly beneficial in financial forecasting, where both historical and forward-looking trends matter.

The BiLSTM model, due to its dual-directional memory capabilities, has emerged as a preferred approach for sequential data in complex environments. Hamayel and Owda (2021) demonstrated that BiLSTM models, in comparison to standard LSTM, deliver superior results in predicting volatile asset prices, such as cryptocurrency, by leveraging both past and future dependencies within the time series. In contexts like stock movement prediction in the European banking system, BiLSTM's ability to consider comprehensive temporal trends may offer an edge, capturing bidirectional dependencies that unidirectional models often overlook. This capability is further supported by Suebsombut et al. (2021), who highlighted the BiLSTM's advantages in time-series data predictions due to its bidirectional structure, showing notable improvements in predictive accuracy for datasets that require modelling of both past and future information.

The cyclic nature of financial stability within the Eurozone significantly affects stock movement predictions for European banks. Bouheni and Hasnaoui (2017) observed that Eurozone banks exhibit procyclical stability behaviours, taking on more risk during economic expansions and tightening during downturns, which impacts the overall volatility in the banking sector. The impact of financial crises on bank stock returns further complicates forecasting within the Eurozone. Allegret et al. (2016) analyzed the Eurozone sovereign debt crisis and demonstrated that this event led to significant contagion among European banks, heavily affecting stock returns across the region. Such economic stressors underscore the importance of incorporating economic indicators related to sovereign risk and macroeconomic stability into forecasting models.

The European banking sector has pursued greater integration, though barriers remain due to legal and economic differences among countries. Kolia and Papadopoulos (2022) examined efficiency convergence within the EU and the Eurozone, noting that while there are signs of convergence, significant disparities persist across countries, affecting banking efficiency and stock volatility. In addition, Apergis et al. (2015) explored the bank lending channel as a function of the European Central Bank's monetary policy, revealing that bank characteristics like stability and size influence how banks respond to monetary policy changes.

The effectiveness of predictive models such as Logistic Regression, XGBoost, LSTM, and BiLSTM is significantly influenced by the size of the training dataset. For Logistic Regression, a minimum of 10 events per predictor variable is recommended to ensure reliable estimates (Peduzzi et al., 1996). XGBoost, a gradient boosting algorithm, can handle smaller datasets but benefits from larger datasets to capture complex patterns effectively (Chen & Guestrin, 2016). Deep learning models like LSTM and BiLSTM require substantial amounts of data due to their numerous parameters and capacity to model intricate temporal dependencies; insufficient data can lead to overfitting and poor generalization (Goodfellow et al., 2016). Therefore, while traditional models like Logistic Regression may perform adequately with smaller datasets, advanced models such as LSTM and BiLSTM necessitate larger datasets to achieve optimal performance**.**

3. Methodology

The objective of this study is to evaluate whether BiLSTM and LSTM can outperform XGBoost and Logistic Regression in predicting the direction of stock prices within the Eurozone banking sector. The task involves predicting the daily stock price movement (up: 1 or down: 0) of the ten largest Eurozone banks by market capitalization. These banks are included in the STOXX600 index, and their importance to the regional financial system makes them an ideal focus for this research.

The dataset used in this study was extracted from Datastream by Refinitiv. Daily closing prices for the ten largest Eurozone banks, as determined by market capitalization, were retrieved alongside twelve independent variables representing financial and macroeconomic indicators. The dataset spans a time frame from January 1, 2000, to May 31, 2024, offering a rich, multi-decade perspective on stock price movements. Predictions were performed on the final 20% of the dataset, corresponding to the most recent observations, allowing for robust out-of-sample evaluation.

In Table 1, the descriptive statistics of the selected bank are presented.

Table 1

Descriptive statistics of selected banks

Bank name	Count	Mean	Std.	Min	25%	50%	75%	Max
BNP PARIBAS	6370	51.79	11.8	20.78	44.48	51.19	57.61	91.6
BANCO SANTANDER	6370	4.82	1.43	1.47	3.7	4.8	5.88	8.4
UNICREDIT	6370	60.58	52.99	6.21	15.1	29.94	108.09	198.41
INTESA SANPAOLO	6370	2.71	1.01	0.87	2.04	2.46	3.2	5.83
BBV ARGENTARIA	6370	7.78	2.74	2.16	5.62	7.31	9.57	15.26

Bank name	Count	Mean	Std.	Min	25%	50%	75%	Max
ING GROEP	6370	13.66	6.51	1.92	8.98	12.24	16.59	33.76
CREDIT AGRICOLE	6370	13.63	5.7	2.88	10.18	12.37	15.16	32.72
ERSTE GROUP BANK	6370	28.2	10.75	6.56	19.54	28.31	35.26	57.63
KBC GROUP	6370	51.57	20.58	5.5	35.67	50.96	65.5	106.24
SOCIETE GENERALE	6370	46.3	24.68	10.9	27.43	42.56	56.75	140.55

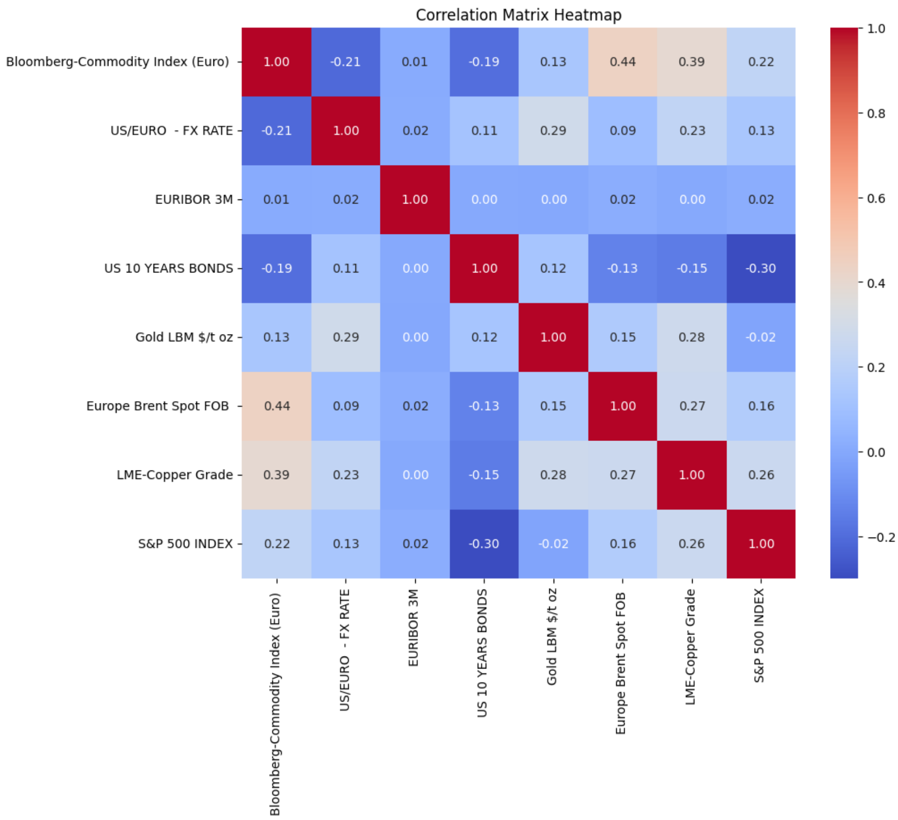
Source: Author's contribution

Extensive data preprocessing was undertaken to ensure the quality and usability of the dataset. Missing values were identified and imputed using the K-Nearest Neighbors (KNN) algorithm, a technique known for its effectiveness in maintaining statistical relationships within the data. Outlier detection was also performed, but no significant anomalies required removal. The twelve initial independent variables underwent a correlation analysis using Pearson's correlation coefficient to identify potential multicollinearity. Variables with a correlation greater than 0.5 were deemed redundant and removed, resulting in a final set of eight independent variables. The variables included in the dataset are: Bloomberg-Commodity Index (Euro), US/EURO - FX Rate, EURIBOR 3M, US 10 Years Bonds, Gold LBM $/t oz, Europe Brent Spot FOB, LME-Copper Grade, and S&P 500 Index.

The Correlation matrix of the independent variables is presented in Figure 1.

Figure 1

Correlation Matrix Heatmap



Source: Author's contribution

To enhance the performance of the models, all features (both dependent and independent variables) were transformed by calculating logarithmic returns. The dataset, comprising 6370 daily observations, was divided into training and testing subsets using stratified sampling. The training set consisted of 80% of the data, while the remaining 20% was reserved for testing. This stratification preserved the class proportions and ensured reliable performance comparisons across models. The dependent variable, representing the banks' stock prices, was derived from the calculated logarithmic returns.

A binary classification framework was adopted, where the target variable was assigned a value of 1 if the logarithmic return was positive, indicating an increase in the stock price. Conversely, a value of 0 was assigned if the logarithmic return was negative, signifying a decrease in the stock price. This approach facilitates a clear distinction between upward and downward price movements, enabling the development of predictive models for directional price changes.

Figure 2 presents the distribution of the target variable.

Figure 2

Distribution of the target variable

Source: Author's contribution

Four different models were implemented to classify stock price direction: Logistic Regression, BiLSTM, LSTM, and XGBoost. Logistic Regression was chosen as the baseline model due to its simplicity and interpretability, offering a benchmark against which the performance of more complex models could be measured. BiLSTM was selected as a state-of-the-art neural network model capable of processing sequential data bidirectionally, enabling it to capture temporal dependencies in both forward and backward directions. LSTM, a foundational recurrent neural network, was included to provide a counterpart to BiLSTM by focusing solely on forward dependencies. Finally, XGBoost, a robust ensemble learning method, was selected for its established performance in classification tasks, particularly with structured datasets.

Logistic Regression was formally introduced by Cox (1958) as a method for regression analysis of binary outcomes and has since become a foundational tool in statistical modeling. Logistic Regression served as the baseline method for the binary classification task. The model was trained using the liblinear solver, which is well-suited for small to medium-sized datasets. The features were scaled using MinMaxScaler to normalize the data and ensure the stability of the logistic model coefficients. The binary cross-entropy loss function was minimized, and accuracy was evaluated as the performance metric. The model was trained on 80% of the data and validated on the remaining 20%, with predictions binarized using a 0.5 threshold.

The BiLSTM methodology was initially introduced by Schuster and Paliwal (1997), who demonstrated its ability to capture bidirectional dependencies in sequential data, and further developed by Graves and Schmidhuber (2005) to improve its application in time-series tasks. The BiLSTM model was designed to leverage the sequential nature of financial data by analysing dependencies in both forward and backward directions. The architecture consisted of two Bidirectional LSTM layers with 64 and 32 units, respectively, followed by dropout layers (20%) to mitigate overfitting. A dense layer with 10 units and a ReLU activation function provided intermediate processing, and a final dense layer with a sigmoid activation function generated the binary output. The model was compiled with the Adam optimizer (learning rate: 0.001), binary cross-entropy as the loss function, and accuracy as the evaluation metric. The training was conducted over 20 epochs with a batch size of 32, incorporating early stopping to prevent overfitting and reduce computational cost. A learning rate scheduler adjusted the learning rate dynamically if the validation loss plateaued for three consecutive epochs.

The LSTM architecture was originally proposed by Hochreiter and Schmidhuber (1997), who addressed the vanishing gradient problem inherent in recurrent neural networks, and was later refined by Gers et al. (2000) to include mechanisms for learning to forget irrelevant information.The LSTM model was configured similarly to the BiLSTM, but it focused solely on forward sequential dependencies. The architecture included two LSTM layers with 64 and 32 units, accompanied by dropout layers to prevent overfitting. The output layer utilized a sigmoid activation function to handle the binary classification problem. The Adam optimizer with an initial learning rate of 0.001 was employed, and binary cross entropy was used as the loss function. Training was performed over 20 epochs with a batch size of 32. Early stopping and learning rate scheduling were employed to enhance training efficiency and avoid overfitting.

XGBoost was introduced by Chen and Guestrin (2016) as a scalable tree-boosting system, building on the gradient-boosting machine framework originally proposed by Friedman (2001). The XGBoost model was employed to evaluate the performance of gradient-boosted decision trees in predicting stock price direction. The model was calibrated with 100 estimators, a maximum tree depth of 6, a learning rate of 0.1, and subsampling and column sampling rates of 0.8 to balance model complexity and generalizability. The binary logistic loss was used as the objective function and the log loss metric guided model optimization. The model was trained on the 80% training data split and tested on the remaining 20%. The predictions were thresholded at 0.5 to assign binary labels.

The evaluation of model performance was conducted using several key metrics. The confusion matrix provided a detailed breakdown of true positives, false positives, true negatives, and false negatives, offering a granular view of classification performance. Area Under the Curve (AUC) was calculated to assess the model's ability to distinguish between binary classes, providing a robust measure of classification quality. Additionally, accuracy was computed as a straightforward indicator of overall predictive performance. These metrics enabled a comprehensive comparison of the models, highlighting the strengths and weaknesses of each approach.

All analyses were implemented in Python, utilizing libraries such as TensorFlow/Keras for neural network models, Scikit-learn for Logistic Regression and evaluation metrics, and XGBoost for gradient-boosted decision trees.

4. Results

Based on the results presented in Table 2, the Logistic Regression model demonstrated the highest average performance among the four models, with a mean AUC of 0.679 and accuracy of 0.634. This indicates that Logistic Regression achieved the best balance between true positive rates and false positive rates across the banks. XGBoost followed closely, with an average AUC of 0.669 and accuracy of 0.617, showing competitive performance but slightly lower predictive ability compared to Logistic Regression.

Table 2

Comparative Analysis of Models

No.	Variable	Logistic Regression		XGBoost		LSTM		BiLSTM	
		AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
1	BNP PARIBAS	0.708	0.652	0.693	0.625	0.709	0.474	0.708	0.474
2	BANCO SANTANDER	0.709	0.657	0.690	0.633	0.707	0.482	0.707	0.482
3	UNICREDIT	0.684	0.641	0.661	0.622	0.684	0.469	0.683	0.469
4	INTESA SANPAOLO	0.674	0.634	0.656	0.602	0.673	0.483	0.674	0.483
5	BBV.ARGENTARIA	0.690	0.644	0.675	0.618	0.689	0.488	0.688	0.488
6	ING GROEP	0.697	0.651	0.691	0.633	0.695	0.469	0.695	0.469
7	CREDIT AGRICOLE	0.655	0.619	0.674	0.625	0.653	0.432	0.649	0.432
8	ERSTE GROUP BANK	0.616	0.591	0.606	0.569	0.614	0.465	0.616	0.465
9	KBC GROUP	0.659	0.606	0.644	0.597	0.657	0.493	0.657	0.493
10	SOCIETE GENERALE	0.692	0.644	0.702	0.644	0.690	0.481	0.692	0.481
Mean		**0.679**	**0.634**	**0.669**	**0.617**	**0.677**	**0.474**	**0.677**	**0.474**

Source: Author's contribution

The LSTM and BiLSTM models exhibited similar mean AUC values (0.677 each) but significantly lower accuracy (0.474 for both). This suggests that while these models performed comparably in distinguishing between positive and negative returns, their overall predictive accuracy was less reliable. Notably, the deep learning models (LSTM and BiLSTM) showed consistent underperformance in accuracy, indicating potential challenges in learning from the dataset or issues with overfitting.

Across individual banks, BNP Paribas, Banco Santander, and ING Groep consistently showed higher AUC and accuracy scores across all models, indicating better predictability for these stocks. Conversely, Erste Group Bank and Credit Agricole showed the lowest AUC and accuracy values, reflecting comparatively weaker model performance for these banks. These results highlight the varying effectiveness of different models depending on the bank and suggest that traditional machine learning methods, such as Logistic Regression and XGBoost, may be better suited for this particular dataset compared to deep learning approaches.

5. Conclusion

This study investigated the predictive capabilities of four models—Logistic Regression, XGBoost, LSTM, and BiLSTM—in forecasting the directional movement of stock prices for the ten largest Eurozone banks by market capitalization. Utilizing a comprehensive dataset spanning from January 1, 2000, to May 31, 2024, the paper incorporated eight financial and macroeconomic indicators to enhance the robustness of the predictions. By focusing on the final 20% of the dataset for out-of-sample evaluation, the study aimed to simulate real-world predictive scenarios and assess the models' practical applicability.

The findings indicate that Logistic Regression outperformed the other models, achieving the highest average AUC of 0.679 and an accuracy of 63.4%. XGBoost followed closely with an average AUC of 0.669 and an accuracy of 61.7%. These results suggest that traditional machine learning models are more effective in this context than advanced deep learning models like LSTM and BiLSTM, which both recorded lower accuracies of 47.4% despite comparable AUC values.

The superior performance of Logistic Regression may be attributed to its simplicity and ability to generalize well with the available data, capturing the essential relationships without overfitting. XGBoost's competitive performance underscores its strength in handling structured data and its robustness against overfitting through regularization techniques. On the other hand, the deep learning models may have underperformed due to the limited dataset size relative to the requirements of such models, leading to challenges in learning complex temporal patterns inherent in financial time series data.

Future research could explore several avenues to enhance predictive accuracy. Incorporating additional data sources such as real-time news feeds, sentiment analysis from social media, or alternative financial indicators might provide deeper insights into market movements. Employing advanced feature engineering techniques and dimensionality reduction methods could help in extracting more relevant features from the existing data. Furthermore, experimenting with hybrid models that combine the strengths of traditional machine learning and deep learning approaches might yield better performance. Adjusting the architectures of LSTM and BiLSTM models, perhaps by integrating attention mechanisms or using transfer learning with larger datasets, could also address the underperformance observed in this study.

In conclusion, while deep learning models hold theoretical appeal for capturing intricate patterns in sequential data, traditional models like Logistic Regression and XGBoost demonstrated more reliable performance in predicting stock price direction in the Eurozone banking sector.

References

1. Allegret, J.-P., Raymond, H., & Rharrabti, H., (2016). The Impact of the Eurozone Crisis on European Banks Stocks: Contagion or Interdependence? European Research Studies Journal, 19(1), pp. 129–148.
2. Apergis, N., Miller, S. M., & Alevizopoulou, E., (2015). The bank lending channel and monetary policy rules for Eurozone banks: further extensions. BE Journal of Macroeconomics, 15(1), pp. 93–112.
3. Chen, T., & Guestrin, C., (2016) XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.
4. Cox, D. R. (1958). The regression analysis of binary sequences. Journal of the Royal Statistical Society: Series B (Methodological), 20(2), pp. 215–232.
5. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), pp. 1189–1232.
6. Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. Neural Computation, 12(10), pp. 2451–2471.

7. Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks, 18(5–6), pp. 602–610.
8. Hamayel, M. J., & Owda, A. Y., (2021). A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms. AI, 2, pp. 477–496.
9. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), pp. 1735–1780.
10. Kolia, D. L., & Papadopoulos, S., (2022). Integration in banking efficiency: a comparative analysis of the European Union, the Eurozone, and the United States banks. Journal of Capital Markets Studies, 6(1), pp. 48–70.
11. Kwak, G.-H., Park, C.-W., Ahn, H.-Y., Na, S.-I., Lee, K.-D., & Park, N.-W., (2020). Potential of Bidirectional Long Short-Term Memory Networks for Crop Classification with Multitemporal Remote Sensing Images. Korean Journal of Remote Sensing, 36(4), pp. 515–525.
12. Qiu, M., & Song, Y., (2016). Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. PLOS ONE, 11(5)
13. Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), 2673–2681.
14. Suebsombut, P., Sekhari, A., Sureephong, P., Belhi, A., & Bouras, A., (2021). Field Data Forecasting Using LSTM and Bi-LSTM Approaches. Applied Sciences, 11, pp. 11820.
15. Zhong, X., & Enke, D., (2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms. Financial Innovation, 5, pp. 24.