World Journal of Information Technology
Print ISSN: 2959-9903
Online ISSN: 2959-9911
DOI: https://doi.org/10.61784/wjit3074

TIME-SERIES FORECASTING OF STOCK PRICE VIA BIDIRECTIONAL LSTM-ATTENTION NEURAL ARCHITECTURE

MingXi Ma
School of Statistics and Data Science, Shanghai University of International Business and Economics, Songjiang 201600, Shanghai, China.

Abstract: Predicting stock prices with high accuracy continues to be a major challenge in financial markets, primarily because of the intricate, non-linear, and highly volatile characteristics of price movements. Traditional statistical methods and standard long short-term memory (LSTM) networks exhibit limitations in capturing temporal dependencies and identifying critical features that significantly influence price movements. To address these challenges, this paper proposes a novel bidirectional LSTM with attention mechanism (BiLSTM-Attention) model for stock price prediction. The proposed model employs bidirectional LSTM layers to process time-series data in both forward and reverse directions concurrently, thereby capturing a more complete picture of past and potential future trends. Additionally, a self-attention mechanism is incorporated to dynamically allocate weights across time steps, enabling the model to focus on salient features that exert substantial influence on price fluctuations. Experimental validation is conducted using real-world stock price data from American International Group (AIG). Results demonstrate that the proposed BiLSTM-Attention model significantly outperforms baseline models across all evaluation metrics, validating the effectiveness of combining bidirectional processing with attention mechanisms for stock price forecasting. The proposed approach offers a stable and effective method for predicting stock prices in the short term.
Keywords: Stock price prediction; Bidirectional LSTM; Attention mechanism; Deep learning; Time series forecasting

1 INTRODUCTION

With the rapid advancement of artificial intelligence and financial technology (fintech), Leveraging big data analysis together with machine learning methods to improve financial market prediction has become an important research focus in both academic and industrial fields [1-2]. As an integral component of the global financial system, stock markets exhibit price dynamics that are shaped by a multitude of interrelated factors, including macroeconomic conditions, investor sentiment, political developments, and corporate performance. These complex interactions render stock markets inherently characterized by high volatility, nonlinearity, and uncertainty [3]. Consequently, accurate stock price forecasting not only provides a scientific foundation for informed investment decision-making but also offers essential guidance for risk management and asset allocation strategies, thereby holding substantial theoretical significance and practical value [4].

Traditional stock price forecasting has predominantly relied on statistical modeling approaches, such as time series analysis and regression models, which were extensively employed in early research. However, as financial data has grown increasingly voluminous and complex, characterized by pronounced nonlinear dynamics, the limitations of these conventional methods have become evident. Specifically, their capacity to process high-dimensional, multi-source data remains constrained, resulting in prediction accuracy that often falls short of practical requirements [5].

In recent years, advances in deep learning technology have provided new avenues for stock price prediction research. Several hybrid architectures have demonstrated promising results by combining complementary neural network structures. For example, the CNN-LSTM framework introduced in employs convolutional neural networks to obtain localized spatial patterns from financial data [6], and then utilizes long short-term memory units to learn extended temporal relationships. Experimental results indicate that this hybrid architecture outperforms standalone LSTM or CNN models. Likewise [7], adopts a bidirectional sliding long short-term memory (BiSLSTM) architecture capable of handling sequences in both the forward and reverse directions, thereby simultaneously exploiting historical and prospective contextual information to achieve more comprehensive temporal modeling and enhanced feature learning capabilities. However, research has shown that LSTM-based models exhibit sensitivity to input data format, quality, and preprocessing strategies [8]. Despite notable progress in deep learning-based approaches, several critical challenges persist, including the need to enhance model prediction accuracy, improve generalization capability across diverse market conditions, and increase model interpretability—issues that remain at the forefront of current research efforts [9].

Motivated by the aforementioned limitations, this paper proposes a novel prediction model that integrates bidirectional long short-term memory (BiLSTM) networks with attention mechanisms (BiLSTM-Attention) to address two primary challenges inherent in standard LSTM architectures for stock price forecasting: high computational complexity and inadequate identification of critical temporal features. The proposed model employs a bidirectional LSTM framework that processes temporal sequences in both forward and backward directions, thereby enabling more comprehensive modeling of complex price dynamics and temporal dependencies. Concurrently, An attention mechanism is integrated to dynamically assign importance across various time steps, enabling the model to prioritize key temporal features that significantly affect price fluctuations. This dual architecture not only expands the model's ability to identify subtle or implicit patterns but also enhances its interpretability and resilience. Experimental results indicate that the BiLSTM-Attention framework effectively models multi-scale temporal dependencies and maintains strong generalization capability under diverse market scenarios, thus offering a stable and efficient approach for short-term financial price prediction.

The remainder of this paper is organized as follows. Section 2 presents the methodology and theoretical framework of the proposed deep learning model. Section 3 describes the experimental design and presents empirical results based on real-world stock price data from American International Group (AIG), where the model's performance is assessed through several evaluation indicators, such as root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2). A comparative evaluation against several baseline models is also performed to highlight the advantages of the proposed method. Section 4 concludes the study by summarizing the main results, outlining existing limitations, and suggesting possible avenues for future investigation.

2 BILSTM-ATTENTION MODEL

2.1 BiLSTM-Attention Layers

Figure 1 illustrates the overall architecture of the proposed BiLSTM-Attention stock prediction model. The model comprises an input layer, a BiLSTM layer, a self-attention layer, a flattening layer, a fully connected layer, and an output layer.

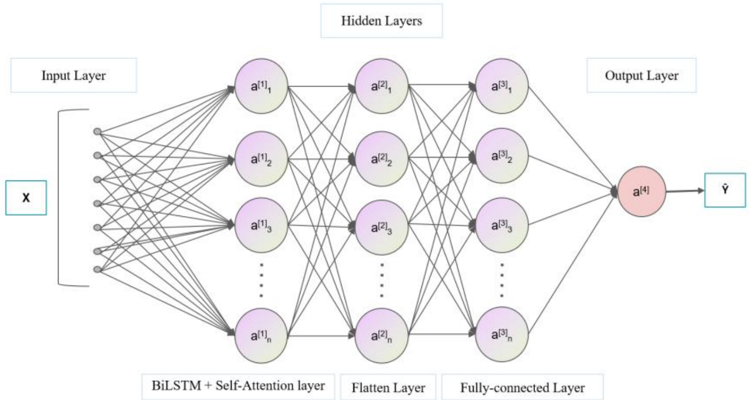

Figure 1 The Architecture of BiLSTM-Attention Model

The input layer receives the stock price sequence samples and passes them to the BiLSTM layer for processing. As shown in Figure 2, the BiLSTM layer consists of forward and backward LSTM units, performing bidirectional processing on the input sequence. The LSTM unit regulates information flow through gating mechanisms, including the forget gate, input gate, candidate memory cell, and output gate. Its internal computations are defined by the following equations:

$$f_t=\sigma(W_f\cdot[h_{t-1},x_t]+b_f), \tag{1}$$

$$i_t=\sigma(W_i\cdot[h_{t-1},x_t]+b_i), \tag{2}$$

$$\widetilde{C}_t=tanh(W_c\cdot[h_{t-1},x_t]+b_c), \tag{3}$$

$$o_t=\sigma(W_o\cdot[h_{t-1},x_t]+b_o), \tag{4}$$

$$C_t=f_t\odot C_{t-1}+i_t\odot\widetilde{C}_t, \tag{5}$$

$$h_t=o_t\odot tanh(C_t), \tag{6}$$

where: W_f, W_i, W_c, W_o denote weight matrices, and b_f, b_i, b_c, b_o denote bias vectors. These parameters are iteratively optimized during training through the backpropagation algorithm. $\sigma(\eta)=\frac{1}{1+e^{-\eta}}$ denotes the Sigmoid activation function, while $tanh(\eta)=\frac{e^{\eta}-e^{-\eta}}{e^{\eta}+e^{-\eta}}$ represents the hyperbolic tangent function, and $\odot$ denotes the dot product operation.

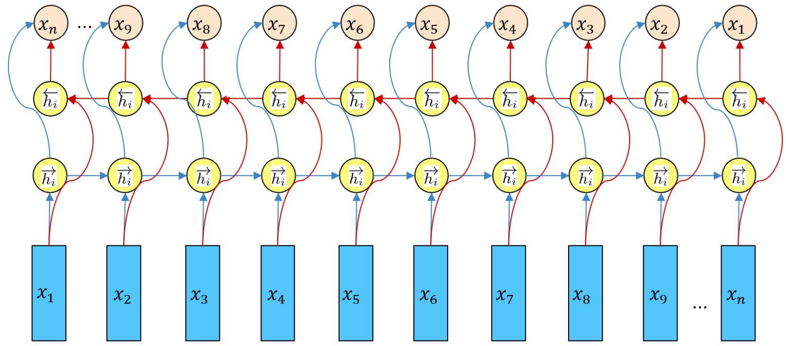

Figure 2 The Architecture of BiLSTM Layer

The forward LSTM's first input is x_1, and its last input is x_n; the backward LSTM processes in reverse order from x_n to x_1. This bidirectional processing enables the network to simultaneously capture both past and future contextual information at the current time step. The output $\boldsymbol{h}_t$ of the BiLSTM at time step t is obtained by concatenating the hidden state $\overrightarrow{\boldsymbol{h}}_t$ of the forward LSTM and the hidden state $\overleftarrow{\boldsymbol{h}}_t$ of the backward LSTM, calculated as:

$$\overrightarrow{\boldsymbol{h}}_t = LSTM_f(x_t, \overrightarrow{\mathrm{h}}_{t-1}), \tag{7}$$

$$\overleftarrow{\boldsymbol{h}}_t = LSTM_b(x_t, \overleftarrow{\mathrm{h}}_{t-1}), \tag{8}$$

$$\boldsymbol{h}_t = [\overrightarrow{\mathrm{h}}_t; \overleftarrow{\mathrm{h}}_t] \tag{9}$$

Through these mechanisms, the BiLSTM layer maintains memory across varying time intervals, capturing temporal dependencies within the input sequence. This bidirectional processing strengthens the network's capability to identify useful features within stock price sequences, which in turn contributes to higher predictive accuracy.

Following the BiLSTM layer, a self-attention module is incorporated to further strengthen the model's capability to extract prominent features. The self-attention mechanism allows the model to reference information across the whole input sequence when computing each position, thus enabling it to capture long-range dependencies in stock price data. As shown in Figure 3, the self-attention layer operates on the BiLSTM output sequence, computing the query matrix $\boldsymbol{Q}$, key matrix $\boldsymbol{K}$, and value matrix $\boldsymbol{V}$:

$$\boldsymbol{Q} = \mathbf{W_q} \cdot \mathbf{H}, \tag{10}$$

$$\mathbf{K} = \mathbf{W_k} \cdot \mathbf{H}, \tag{11}$$

$$\mathbf{V} = \mathbf{W_v} \cdot \mathbf{H}, \tag{12}$$

$$\mathbf{Z} = \mathbf{Attention(Q,K,V)} \cdot \mathbf{V} = softmax\left(\frac{\mathbf{QK^T}}{\sqrt{d_k}}\right) \cdot \mathbf{V}, \tag{13}$$

where $\boldsymbol{W_q}$, $\boldsymbol{W_k}$ and $\boldsymbol{W_v}$ are learnable weight matrices, and d_k denotes the dimension of the attention mechanism. The Softmax function is applied row-wise, generating a probability distribution for each query at positions within the input sequence.

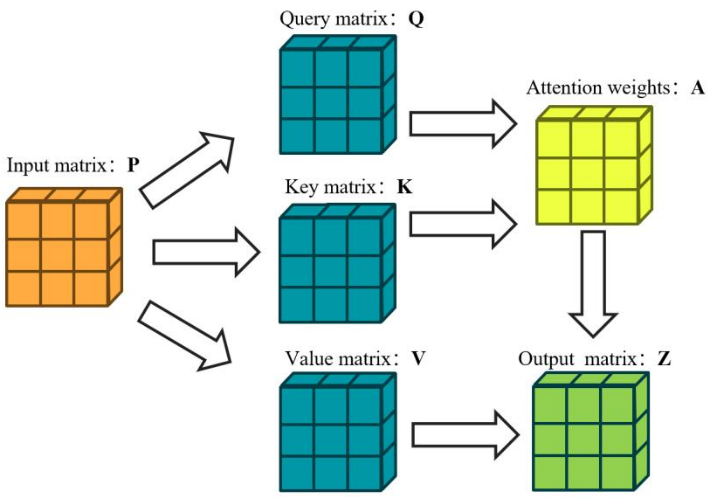

Figure 3 The Architecture of Self-Attention Layer

The self-attention layer directly models the relationship between any two positions in the sequence, which is particularly crucial for capturing complex nonlinear dynamics in stock price data. By providing a global view of the sequence, the self-attention layer complements the local patterns captured by the BiLSTM layer, enabling a feature extraction process that leverages both local and global information. This synergy enhances the model's ability to distinguish informative features from noise, thereby improving prediction robustness.

The feature matrix $\mathbf{Z}\in\mathbb{R}^{T\times 2d}$ output by the Self-Attention layer is subsequently fed into a Flatten Layer. This operation reshapes the two-dimensional feature matrix into a one-dimensional vector $z_{flat}\in\mathbb{R}^{T\times 2d}$, providing a standardized input for subsequent fully connected layers. Mathematically, this operation can be expressed as:

$$z_{flat}=Flatten(\mathbf{Z}). \tag{14}$$

This operation preserves feature information across all time steps while converting it into a format suitable for the fully connected layer. The flattened feature vector is then fed into the fully connected layer, where nonlinear transformations integrate the learned features to generate the regression output. The output $\hat{y}$ from the fully-connected layer represents the predicted stock price, calculated as:

$$\hat{y}=\mathbf{W}_{fc}\cdot\ z_{flat}+b_{fc}, \tag{15}$$

where $\mathbf{W}_{fc}\in\mathbb{R}^{(T\times 2d)\times 1}$ denotes the weight matrix, and $\mathbf{W}_{fc}\in\mathbb{R}^{(T\times 2d)\times 1}$ represents the bias term.

Through end-to-end training, the entire architecture can automatically learn complex nonlinear mapping relationships from raw price sequences to predicted outputs, enabling accurate stock price forecasting.

2.2 BILSTM-ATTENTION Training

During offline training, the model parameters are iteratively updated on the training set. As training progresses, the discrepancy between predicted and actual values gradually diminishes and converges. Specifically, let $Y=\{y_1,\ y_2,\ \ldots,\ y_N\}$ denote the actual stock price, $X=\{x_1,\ x_2,\ \ldots,\ x_N\}$ denote the corresponding input price feature vector, and $\hat{Y}=\{\hat{y}_1,\ \hat{y}_2,\ \ldots,\ \hat{y}_N\}$ denote the model's predicted price. The training objective is to minimize the loss function, which quantifies the discrepancy between actual prices and predicted prices . This study employs the mean squared error (MSE) as the loss function:

$$\Gamma=\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2 \tag{16}$$

where denotes the sample size. The model calculates gradients of the loss function with respect to all parameters using backpropagation and updates the network weights using the Adam optimizer. Through this iterative process, the model learns the mapping from input features to target prices, thereby minimizing prediction error.

3 EXPERIMENTAL SIMULATION

To validate the effectiveness of the proposed BiLSTM-Attention model, experiments are conducted using stock market data. The model performance is quantitatively compared with baseline methods using multiple evaluation metrics, including RMSE.

3.1 Experimental Dataset

This study uses stock market data from the publicly available dataset on GitHub https://github.com/Deamoner/ultimate-stock-machine-learning-training-dataset/tree/master/full_history. The dataset comprises historical trading data from major publicly traded companies, providing comprehensive coverage across industry sectors and market conditions. The experiments focus on AIG (American International Group) stock data spanning from 1973 to 2019. The dataset includes daily open, high, low, close (OHLC) prices and trading volume. Table 1 presents a sample of the data.

Table 1 Time Series Data Samples

date	volume	open	close	high	low	adjclose
2019-04-16	4709600	46.05	46.74	46.89	46.04	46.74
2019-04-17	3650600	46.75	45.97	46.80	45.62	45.97
2019-04-18	3729300	45.90	46.04	46.44	45.81	46.04

3.2 Evaluation Indicators

To thoroughly assess the forecasting capabilities of various models, this study utilizes four widely recognized statistical indicators: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2). The corresponding formulas are presented as follows [10]:

$$\mathrm{RMSE}=\sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2}, \tag{17}$$

$$\mathrm{MAE}=\frac{1}{m}\sum_{i=1}^{m}|y_i-\hat{y}_i|, \tag{18}$$

$$\mathrm{MAPE}=\frac{100\%}{m}\sum_{i=1}^{m}\left|\frac{y_i-\hat{y}_i}{y_i}\right|, \tag{19}$$

$$\mathrm{R}^2=1-\frac{\sum_{i=1}^{m}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{m}(y_i-\bar{y})^2}. \tag{20}$$

3.3 Experimental Process

The dataset is divided into training and testing subsets with an 80:20 split. Time-series samples are generated using a sliding window method, where each window spans T trading days. The resulting data are then formatted into a three-dimensional tensor (N, T, F) suitable for LSTM input, with N representing the total number of samples, T the sequence length, and F the number of features.

Four models are evaluated: LSTM, BiLSTM, LSTM-Attention, and BiLSTM-Attention. The models are trained using the Adam optimizer with the MSE loss function and a learning rate scheduler to facilitate convergence. Model performance is evaluated using RMSE, MAE, MAPE, and the coefficient of determination (R^2). Table 2 summarizes the key hyperparameter settings.

Table 2 Primary Hyperparameter Settings for the Model

Parameter Name	Parameter value	explanation
Time step	60	Predict the next day based on the past 60 days
Learning rate	0.003	Adam Optimizer Initial Learning Rate
Batch size	64	Number of samples per gradient update
Number of training cycles	50	Number of iterations
Training/Testing Ratio	0.8/0.2	Data Allocation Ratio
First LSTM layer	64	Capture long-term trends
Second LSTM layer	32	Capture local features
Attention Layer	Enable	Dynamic Time Weighting
Learning Rate Scheduling	Decreases to 20% of the original value every 75 rounds.	Prevent overfitting

3.4 Experimental Result

Figure 4 compares the predicted prices from the four models against actual prices. The BiLSTM-Attention model effectively tracks the real price movements, especially in volatile periods, indicating its capability to precisely identify points of price reversal.

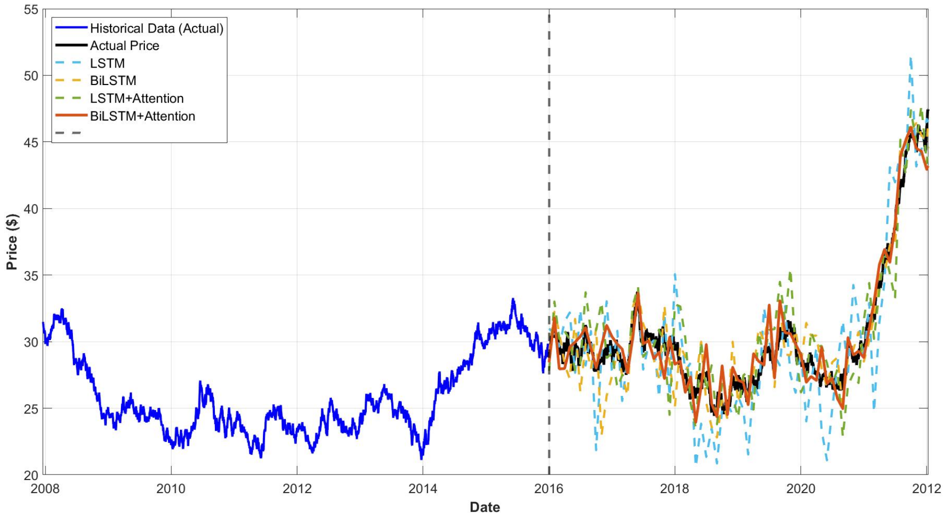

Figure 4 Comparison of the Prediction Curves of the Four Models Against Actual Prices

Figure 5 illustrates the temporal prediction errors of the four models on the test set. Compared to the other three models, the BiLSTM-Attention model displays notably lower error variability, and its smoother error trajectory reflects enhanced stability. Specifically, the prediction errors predominantly lie within ±1 and fluctuate symmetrically around zero, with no significant systematic bias observed, indicating unbiased predictions.

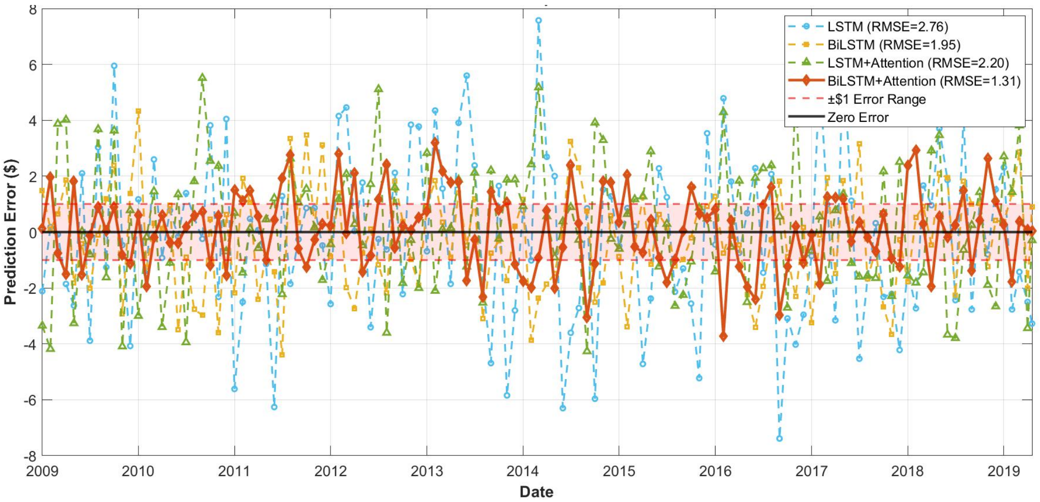

Figure 5 Prediction Error of the Four Models over Time on the Test Set

Table 3 presents the quantitative performance comparison of the four models on the test set. The BiLSTM-Attention model achieves superior performance across all metrics. Compared to the baseline LSTM model, it reduces RMSE by 52.7%, MAE by 64.4%, and MAPE by 63.4%, while improving R^2 from 0.9494 to 0.9887. These results demonstrate that the combination of bidirectional processing and attention mechanism enables the model to effectively extract salient temporal features and focus on critical information, thereby significantly enhancing prediction accuracy.

Table 3 The Results of the Three Models on the Test Set

model	RMSE	MAE	MAPE (%)	R^2
LSTM	2.7635	2.6560	6.01	0.9494
BiLSTM	1.9472	1.5195	3.60	0.9749
LSTM+ Attention	2.2009	1.6968	4.11	0.9679
BiLSTM + Attention	1.3069	0.9449	2.20	0.9887

4 CONCLUSION

This study proposes a BiLSTM-Attention model to address the limitations of standard LSTMs in stock price forecasting, particularly their unidirectional information flow and inability to identify critical time steps. The model combines bidirectional LSTM architecture to capture temporal patterns from both directions with a self-attention mechanism that dynamically weights time steps to focus on key price-influencing moments. Experiments on AIG stock data demonstrate that the BiLSTM-Attention model significantly outperforms baseline models across all metrics, validating the effectiveness of combining bidirectional processing with attention mechanisms. However, the model relies primarily on historical price data and does not incorporate multimodal information such as news sentiment. Future research directions include: (1) integrating text sentiment analysis for multimodal fusion, (2) applying transfer learning to enhance cross-market generalization, (3) extending the model to high-frequency trading and portfolio optimization, and (4) improving interpretability through techniques such as SHAP analysis to provide actionable insights for quantitative investing.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Henrique M B, Sobreiro A V, Kimura H. Literature Review: Machine Learning Techniques Applied to Financial Market Prediction. Expert Systems with Applications, 2019, 124: 226-251.

[2] Zhao Tingting, Han Yajie, Yang Mengnan, et al. A Review of Machine Learning-Based Methods for Time Series Data Forecasting. Journal of Tianjin University of Science and Technology, 2021, 36(5): 1-9.

[3] Mailagaha M K, Christoph L, Pasi L, et al. Machine Learning Techniques and Data for Stock Market Forecasting: A Literature Review. Expert Systems with Applications, 2022, 197.

[4] Xing Weichen. Stock Forecasting in the Big Data Era. China Business Review, 2020(3): 31-32.

[5] Vuong H P, Phu H L, Nguyen V H T, et al. A Bibliometric Literature Review of Stock Price Forecasting: From Statistical Model to Deep Learning Approach. Science Progress, 2024, 107(1).

[6] Bambang S, Derwin S. Utilizing BERT and CNN-LSTM in Stock Price Prediction Using Data Sentiment Analysis and Technical Analysis of Stock and Commodity. Bulletin on Innovative Computing, Information, and Control, 2023, 17(2): 171-176.

[7] Wang Haiyao, Wang Jianxuan, Chen Lihui, et al. A Stock Closing Price Prediction Model Based on CNN-BiSLSTM. Complexity, 2021.
[8] Yadav H, Thorat A. NOA-LSTM: An Efficient LSTM Cell Architecture for Time Series Forecasting. Expert Systems with Applications, 2024, 238.
[9] Yang Zheng, Wu Haocheng, Zhang Jing, et al. Can Dimension Reduction Data ForecastingEnhance the Predictive Performance of Stock Excess Returns?. Journal of Econometrics, 2023, 3(3): 828-847.
[10] Mao Yueyue, Zhang Qiuyue. Research on Machine Learning-Based Stock Forecasting Methods. Modern Computer, 2020(23): 44-47.