GATR-Global Journal of Business and Social Science Review

GATR JOURNALS

Journal homepage: http://gatrenterprise.com/GATRJournals/GJBSSR/vol14_2026_issue1.html

GATR-Global J. Bus. Soc. Sci. Review 14(1) 24 – 30 (2026)

Short-Term Stock Price Prediction Based on Single and Stacking Machine Learning Models

Cross Mark

Wenchuan Sun[1], Chia Yean Lim[1]*, Fengqi Guo[2], Sau Loong Ang[3]

[1]School of Computer Sciences, Universiti Sains Malaysia, 11800, Minden, Malaysia
[2]CITIC Securities, 150000, Harbin, China
[3]Department of Computing and Information Technology, Tunku Abdul Rahman University of Management and Technology, Penang Branch, 11200, Tanjung Bungah, Malaysia

ABSTRACT

Objective: As the investment environment improves, individuals are increasingly eager to invest their idle funds. Securities companies have become the preferred choice for buying financial products. The current accuracy of stock predictions relies on the comprehensive models used by each securities company, including stock market trading, data, and stock pricing models. However, securities companies have not adequately explored a single suitable model for stock predictions and have rarely assessed the effectiveness of stacking and ensemble methods in improving these predictions.
Methodology/Technique – This research first explored and proposed the best single-stock prediction model. Next, it combined four individual prediction models to create a stacking model.
Finding – The comparison between the single and stacking models demonstrated that the stacking model's prediction accuracy exceeded that of the single model. Therefore, it is recommended that securities companies adopt a stacking-type prediction model to forecast share prices for their investment customers.
Novelty – Using a stacking model could improve the accuracy of stock price predictions for investment managers, help users make better decisions, and ultimately enhance the company's earnings by delivering more accurate investment outcomes.

Type of Paper: Empirical
JEL Classification: F17, F47
Keywords: Long short-term memory, random forest model, stacking model, stock prediction, support vector machine, XGBoost model.

Reference to this paper should be referred to as follows: Sun, W; Lim, C.Y; Guo, F; Ang, S.L. (2026). Short-Term Stock Price Prediction Based on Single and Stacking Machine Learning Models, *GATR-Global J. Bus. Soc. Sci. Review,* 14(1), 24–30. https://doi.org/10.35609/gjbssr.2026.14.1(2)

1. Introduction

Traditional investment management relies on investor experience and judgment, limiting the use of modern data analysis and machine learning. The stock market is vital to a country's economy, facilitating capital flows, enhancing liquidity, and improving transparency in corporate governance (Levine, 1997). Mainland China has two major stock exchanges: Shanghai and Shenzhen. Investors can open accounts with more than 140 securities firms, including CITIC and Haitong, to trade equities.

*Paper Info: Revised: January 11, 2026
Accepted: March 31, 2026
*Corresponding author: Chia Yean Lim
E-mail: cylim@usm.my
Affiliation: School of Computer Sciences, Universiti Sains Malaysia, 11800, Minden, Malaysia

These firms analyse stocks using four primary methods: fundamental analysis, technical analysis, macroeconomic analysis, and quantitative models (Bloomberg Intelligence Reports, 2023). Chinese firms relied on outdated machine-learning models for stock predictions, necessitating more advanced techniques to improve the efficiency and accuracy of investment decisions. This research aims to develop and compare single- and stacking-stock prediction models in China. It examined four individual machine learning models for short-term stock predictions and then created a stacking model from them. Their predictive accuracy will be analysed, and the best-performing model will be recommended to Chinese firms to aid client investment decisions. The research hopes to enhance stock purchase decision-making by reducing time and minimising human error.

2. Literature Review

Many stock analysis indicators aid in assessing purchase decisions. The Technical Analysis Library (TA-Lib) is a comprehensive Python library with over 150 technical indicators for stock and futures trading. It includes 10 sub-segments: overlap studies, momentum indicators, volume indicators, cycle indicators, price transformation, volatility indicators, pattern recognition, statistical functions, mathematical transformations, and Mathematical operators. Key indicators include the simple moving average (SMA_20), the exponential moving average (EMA_20), the relative strength index (RSI_14), Bollinger Bands, and moving average convergence/divergence (MACD). These indicators help investors decide when to buy and sell.

Modern machine learning (ML) tools analyse complex market patterns in vast financial data, leading to more informed investment decisions. Several techniques, such as Support Vector Machine (SVM), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Long Short-Term Memory (LSTM), have shown success in investment management.

SVM is a supervised learning algorithm used for classification and regression by finding the best hyperplane to segment data points. In stock prediction, technical indicators are combined to assess future trends. It effectively processes multi-dimensional data and analyses multiple indicators simultaneously. It is insensitive to extreme values and can handle certain extreme situations. In cases of limited stock data, it performs well, particularly with newly listed stocks. This model has predicted the stock market (Nti et al., 2020a) and Tesla's closing price (Madhusudan, 2020). Conversely, (Endri et al., 2020) used SVM to predict the delisting of Islamic stocks (ISSI). RF is an ML algorithm of multiple decision trees. It effectively handles incomplete stock data and mitigates the risk of overfitting in the presence of noise. (Khaidem et al., 2016) applied RF to predict stock market movements, achieving high accuracy and robustness against volatility. (Sadorsky, 2021) used an RF approach to predict clean energy stock prices. (Zheng et al., 2024). Also investigated the RF model's capability in the US stock market.

XGBoost is an ML algorithm based on gradient boosting that uses an approximation algorithm to manage continuous features and employs regularisation to prevent overfitting. As an optimised framework, it has gained popularity in finance for its efficiency and predictive accuracy. The XGBoost model performed well on extensive stock data and accommodates a range of characteristics. (Krauss et al., 2017) demonstrated its effectiveness in statistical arbitrage on the S&P 500, achieving higher returns than traditional methods. XGBoost also outperformed other models, as shown in studies by (Xia et al., 2020) and (Yang et al., 2021). The LSTM model, a type of recurrent neural network (RNN), learns long-term dependencies and decides whether to remember or forget data. This experiment can consider multiple stock factors simultaneously and capture both long-term and short-term dependencies of stock prices. (Abe & Nakayama, 2018) used LSTM for forecasting stock returns, achieving significant improvements over traditional methods by effectively modelling temporal dependencies in market data. In model comparisons, researchers found that LSTM outperformed traditional models such as XGBoost, LightGBM, GBDT, CatBoost, logistic regression, and KNN; Nabipour et al., 2020; Han et al., 2022.

A weighted ensemble model is a powerful predictive tool that incorporates multiple models. It combines multiple fundamental models, assigns weights to each based on their performance, and ultimately produces final predictions by combining them (Ye et al., 2022). For example, SVM, RF, XGBoost, and LSTM models can be combined, with the final prediction computed as a weighted average. This model’s advantage lies in its ability to combine the strengths of multiple models, enhance forecast stability, mitigate the risk of relying on a single model, and deliver more accurate predictions. A stacked ensemble model combines multiple models. The forecasts produced by a centralised machine learning algorithm serve as inputs to a second-layer learning algorithm. This algorithm is trained to refine the combined model's predictions, producing a new set of forecasts. (Gyamerah et al., 2019) investigated stacked ensemble learning methods for predicting stock market trends. They found that a stacked ensemble approach using two basic learners (AdaBoost and KNN) alongside GBM as meta-classifiers outperformed individual classifiers. Conversely, (Nti et al., 2020b) introduced a method that employed three fundamental algorithms: decision tree (DT), SVM, and neural network (NN), to create 25 distinct ensemble classifiers and regressors utilising Stacking, Blending, Bagging, and Boosting. Their findings indicated that stack and hybrid techniques yield the best results for stock market predictions, though they entail high computational costs.

3. Research Methodology

The objective of this research is to propose a machine-learning-based stock prediction model to enable securities companies to achieve more accurate forecasts than traditional methods. To achieve this objective, two models are applied for stock prediction. The first type involved four single ML models: SVM, Random Forest, XGBoost, and LSTM. The second type consisted of an ensemble model that utilised a weighted stacked approach to predict stock prices. The advantage of the ensemble model is that it can capture different characteristics from multiple models and integrate multiple perspectives, thereby reducing the inherent deviation of a single model.

This research adopted the Cross-Industry Standard Process for Data Mining (CRISP-DM) model. During the data understanding phase, Apple stock data comprising 504 observations from 2022 to 2023 was obtained. The data included stock code, trading date, closing price, opening price, daily high, daily low, previous day’s closing price, percentage of price change, transaction amount, and volume-weighted average price. During data preprocessing, several comprehensive technical indicators were derived through feature engineering of the dataset, informed by expert advice from a security firm. The indicators and descriptions are shown in Table 1.

The study divided the stock data into four periods, corresponding to the first and second halves of 2022 and 2023, respectively. Data from the entire year of 2022 and the first half of 2023 comprised the training dataset (75%), while data from the second half of 2023 comprised the test set (25%). This approach ensured that the training set was sufficiently large to capture data features and sufficient to validate model performance.

Figure 1 shows the workflow for the stacking ensemble model. First, four basic models were trained, producing preliminary predictions that were used as new features in a secondary meta-model (Gradient Boosting Regressor (GBR) or Ridge). The GBR was favoured over Ridge due to its superior predictive performance. Following on, meta-learning occurred in the second layer (meta model), with the final output in the third layer.

Table 1. Generated indicators and description

No.	Features	Data Type	Description
1.	SMA_20	Numerical	20-day simple moving average
2.	EMA_20	Numerical	20-day exponential moving average
3.	RSI_14	Numerical	14-day relative strength indicator
4.	upper_band	Numerical	Middle band + 2 standard deviations
5.	middle_band	Numerical	20-day moving average
6.	lower_band	Numerical	Middle band - 2 standard deviations
7.	MACD	Numerical	Fast EMA(12-day) - Slow EMA(26-day)
8.	MACD_signal	Numerical	9-day EMA of MACD
9.	MACD_hist	Numerical	MACD – Signal

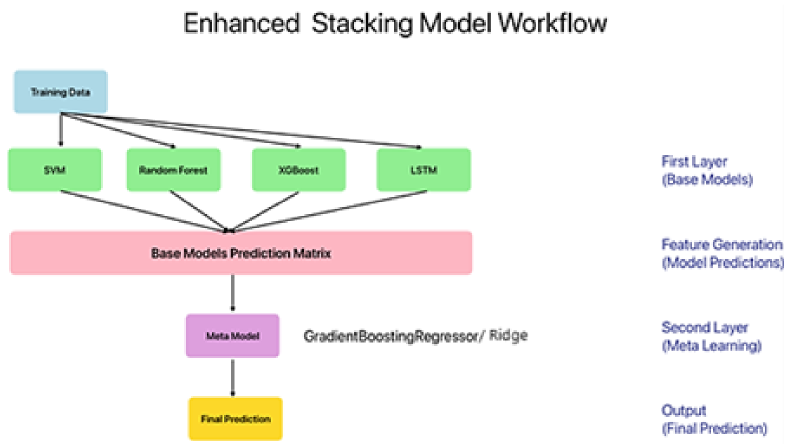

Figure 1. Structure of the proposed stacked model

4. Results

4.1 Single ML Prediction Models

The result of the single model's experiment is summarised in Table 2. Based on observation, SVM is prone to over-prediction. The RF model predicts the long-term trend more accurately, but it needs improvement in handling short-term market fluctuations. The XGBoost model shows some predictive ability, but it still struggles to respond to extreme market conditions. Compared with other models, LSTM shows strong tracking ability throughout the forecast period (especially in the 0-50 and 200-250 periods). The experimental results showed that LSTM achieved the best performance, with the lowest MSE of 19.3742 and RMSE of 4.4916.

Figure 2 shows results from four models. The LSTM model (purple dotted line) closely aligns with the actual results, accurately predicting fluctuations and peaks. In contrast, the SVM model (orange dotted line) produces smooth forecasts that do not track fundamental price changes. The Random Forest and XGBoost models effectively capture medium-term price trends. In conclusion, the LSTM model is the best single-prediction model compared to SVM, RF, and XGBoost.

Table 2 Single Model Experiment Result

Model	Features used	MSE	RMSE	Observation
SVM model	'SMA_20', 'EMA_20', 'RSI_14', 'upper_band', 'lower_band', 'MACD' and 'MACD_signal'	70.8640	8.4181	The model showed intense volatility, especially in the 100-150 period. There is significant over-prediction, with predicted values much higher than actual stock prices.
Random Forest model	'SMA_20', 'EMA_20', 'RSI_14', 'upper_band', 'lower_band', 'MACD' and 'MACD_signal'	54.7088	7.3965	The result showed significant over-forecasting during the 100-150 time period, while actual price movements are accurately tracked in other periods.
XGBoost model	'SMA_20', 'EMA_20', 'RSI_14', 'upper_band', 'lower_band', 'MACD' and 'MACD_signal'	34.3183	5.8582	The results showed that the model tracked well in the early phase (0-50) and captured price fluctuations, but it overpredicted during the 100-150 period. Forecasts after 200 time points are conservative and do not reflect the sharp decline in actual prices.
LSTM model	First input with data of the stock in the first 10 days for time series features extraction	19.3742	4.4016	The result showed that there is some forecasting bias during the 100-150 period, but the overall performance remains stable and accurate.

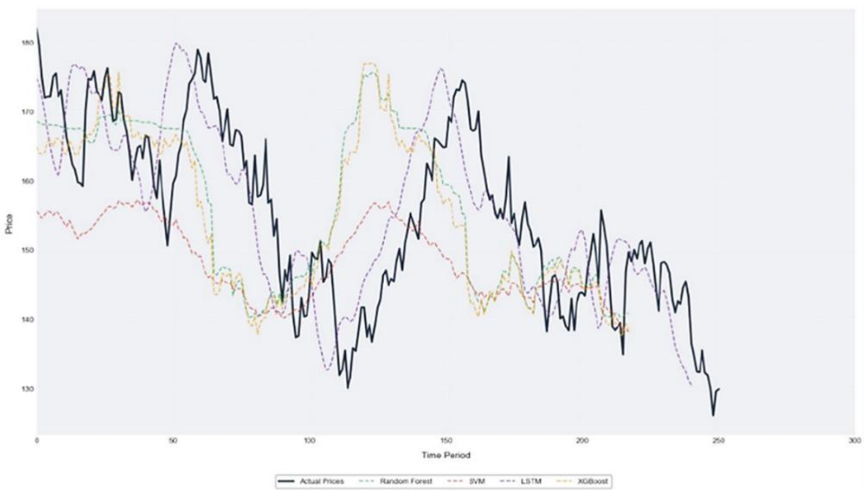

Figure 2 Comparison of the prediction output of all single models

4.2 Ensemble Stacking Model

The stacked model includes individual models such as SVM, random forests, XGBoost, and LSTM. In this setup, predictions from trained models serve as inputs to the next meta-model, which computes optimal weights before making final predictions in the final layer. Ridge and GBR meta-models are compared to determine the best-performing model. GBR outperformed the Ridge metal-model, yielding lower MSE (Rigle: 13.7370, GBR MSE: 0.0466) and average error per share (Rigle: \$3.70, GBR: \$0.22). Figure 3 compares actual prices to predictions from the stacked model using Ridge (purple line) and GBR (red line). The predicted prices from the stacked GBR model closely align with the actual prices (blue line), demonstrating the model’s accuracy.

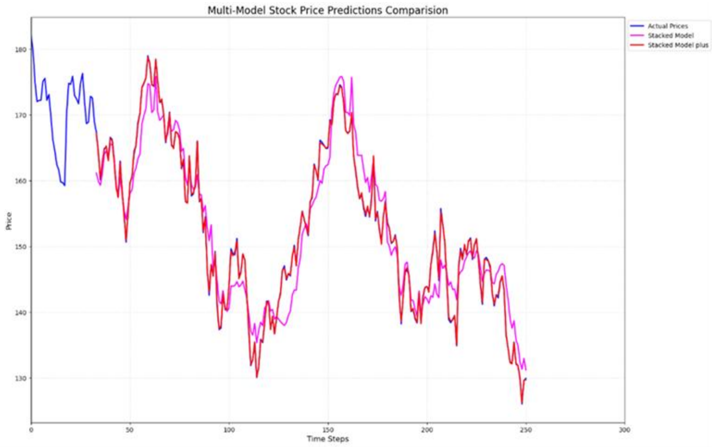

Figure 3. Comparison of actual prices against predictions from the stacked model

4.3 Comparison of the Single and Stacking Prediction Model

The single-model results indicate that LSTM is the top performer. In contrast, the Stacking Model with GBR outperformed the Ridge meta-model. The comparison revealed that stacking provides superior predictions, with MSE and RMSE significantly lower than those of the single model (MSE: 19.3742, RMSE: 4.4016) and the stacking model (MSE: 0.04660, RMSE: 0.2159).

5. Conclusion

This research explores stock prediction using both single- and ensemble-based machine learning models for security firms in China. The single-model experiment found that the LSTM model produced the best predictions among the RF, SVM, and XGBoost models. Therefore, security firms may consider adopting an LSTM model for stock price prediction if they prefer a single model. In contrast, the ensemble model, which comprised four single models - RF, SVM, XGBoost, and LSTM with GBR as a meta-model, outperformed the ensemble model, which used the Rigle meta-model. The ensemble stacking model yielded better predictions than any of the single models studied. As the ensemble model considers more factors and scenarios, security firms can gradually implement it in real business settings to assist their clients in making stock purchase decisions.

References

Abe, M., & Nakayama, H. (2018). Deep learning for forecasting stock returns in the cross-section. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 273–284.

Bloomberg Intelligence Reports. (2023). *Bloomberg Intelligence Reports*.

Endri, E., Kasmir, K., & Syarif, A. (2020). Delisting sharia stock prediction model based on financial information: Support Vector Machine. *Decision Science Letters*, *9*(2), 207–214.

Gyamerah, S. A., Ngare, P., & Ikpe, D. (2019). On stock market movement prediction via stacking ensemble learning method. *2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, 1–8.

Han, D., He, M., & Wang, L. (2022). Bitcoin or gold? A financial investment model based on LSTM. *Frontiers in Business, Economics and Management*, *4*(3), 72–77. https://doi.org/10.54097/fbem.v4i3.1139

Khaidem, L., Saha, S., & Dey, S. R. (2016). Predicting the direction of stock market prices using random forest. *ArXiv Preprint ArXiv:1605.00003*.

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, *259*(2), 689–702.

Levine, R. (1997). Financial development and economic growth: views and agenda. *Journal of Economic Literature*, *35*(2), 688–726.

Madhusudan, D. M. (2020). Stock closing price prediction using machine learning svm model. *International Journal for Research in Applied Science and Engineering Technology*.

Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., Salwana, E., & S, S. (2020). Deep learning for stock market prediction. *Entropy*, *22*(8), 840.

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020a). A comprehensive evaluation of ensemble learning for stock-market prediction. *Journal of Big Data*, *7*(1), 20.

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020b). Efficient stock-market prediction using ensemble support vector machine. *Open Computer Science*, *10*(1), 153–163.

Qiu, J., Wang, B., & Zhou, C. (2020). Forecasting stock prices with long-short term memory neural network based on attention mechanism. *PloS One*, *15*(1), e0227222.

Sadorsky, P. (2021). A random forests approach to predicting clean energy stock prices. *Journal of Risk and Financial Management*, *14*(2), 48.

Wei, X., Ouyang, H., & Liu, M. (2022). Stock index trend prediction based on TabNet feature selection and long short-term memory. *Plos One*, *17*(12), e0269195.

Xia, Z., Xue, S., Wu, L., Sun, J., Chen, Y., & Zhang, R. (2020). ForeXGBoost: passenger car sales prediction based on XGBoost. *Distributed and Parallel Databases*, *38*(3), 713–738.

Yang, Y., Wu, Y., Wang, P., & Jiali, X. (2021). Stock price prediction based on xgboost and lightgbm. *E3s Web of Conferences*, *275*, 1040.

Ye, Z., Wu, Y., Chen, H., Pan, Y., & Jiang, Q. (2022). A stacking ensemble deep learning model for bitcoin price prediction using Twitter comments on bitcoin. *Mathematics*, *10*(8), 1307.

Zheng, J., Xin, D., Cheng, Q., Tian, M., & Yang, L. (2024). The random forest model for analyzing and forecasting the us stock market in the context of smart finance. *ArXiv Preprint ArXiv:2402.17194*.