Stock Prediction Based on Traditional Statistical Models, Machine Learning Models and Fusion Models

Yu Du
School of Data Science, Capital University of Economics and Business, 100000, Beijing, China

Keywords: Stock Prediction, Statistical Models, Machine Learning Models, Fusion Models.

Abstract: This study aims to evaluate how well machine learning (ML) algorithms and classic time series analysis methods can forecast stock market trends. Accurate forecasts of stock prices can greatly aid professionals and investors in making strategic decisions owing to the unpredictable nature of the stock market. This research aims to create a composite model that combines the accuracy of traditional statistical models, which are good at making short-term predictions, with the capabilities of machine learning models that can handle large amounts of complex and nonlinear data. The goal is to enhance the precision of long-term stock price forecasts. This research aims to assess the strengths and weaknesses of four distinct models: Autoregressive Integrated Moving Average(ARIMA), Generalized Autoregressive Conditional Heteroskedasticity(GARCH), Long Short-Term Memory (LSTM), and Random Forest(RF) through training and evaluation with historical stock market data. Additionally, a comparison between these distinct models and an integrated model will be conducted as part of the investigation to develop a more reliable tool for informing investment decisions.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

1. INTRODUCTION

The stock market serves as a barometer for a nation's economic and fiscal dynamics (Lu et al., 2021). Consequently, investors are highly concerned about the future trend of stock values (Li, Pan and Huang, 2019). Nevertheless, numerous variables influence stock prices, including alterations in national policies, fluctuations in the local and global economic landscape, and shifts in the international situation (Sim, Kim and Ahn, 2019). Consequently, forecasting stock prices is a formidable undertaking. Producing accurate predictions of stock prices can greatly reduce the level of risk for investors. These forecasts allow investors to integrate projected stock values into their investing strategy, so increasing the potential for enhanced investment returns (Lu et al., 2021).

In recent times, a wide range of approaches and frameworks have emerged for forecasting stock values. The methodologies can be broadly categorized into two types: classic statistical models and machine learning-based models. When it comes to predicting time series, particularly for short-term predictions, the ARIMA model is widely regarded as more resilient and effective compared to the most used artificial neural network techniques. Other statistical models are generalized autoregressive conditional heteroskedasticity (GARCH) regression and exponential smoothing (Ariyo et al., 2014). However, statistical models ignore the effects of external factors other than the time factor and are all based on the premise that there will be no sudden changes in the market in the future, so statistical models alone are not sufficient for some special cases. Machine learning advancements have facilitated the utilization of ML methods such as random forests and LSTM networks to decipher complex nonlinear patterns in financial datasets. The LSTM network, a variant of a recurrent neural network, is highly proficient in many applications due to its ability to accurately differentiate between recent and past data points by assigning distinct weights and selectively eliminating irrelevant information that is not crucial for future predictions. Unlike other types of recurrent neural networks that primarily handle short-term data sequences, this specific model excels at managing longer input sequences, making it more suitable for applications that require retaining substantial history information. As a result, it is highly effective at predicting stock prices when applied to nonlinear datasets with a huge volume of data (Sunny et al., 2020 & Nelson et al., 2017). Conversely, the

Du, Y.
Stock Prediction Based on Traditional Statistical Models, Machine Learning Models and Fusion Models.
DOI: 10.5220/0013007100004601
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 1st International Conference on Innovations in Applied Mathematics, Physics and Astronomy (IAMPA 2024)*, pages 160-168
ISBN: 978-989-758-722-1
Proceedings Copyright © 2024 by SCITEPRESS – Science and Technology Publications, Lda.

substantial level of noise and frequent fluctuations in crucial characteristics within the stock market render stock prediction intricate and inefficient. Random forests have the ability to conduct feature analysis, which quantifies the significance of each input feature. Utilizing Random Forest (RF) for feature extraction can enhance the precision of stock price forecasts (Ma, Han and Fu, 2019).

This work presents a novel strategy that integrates statistical and machine learning method to overcome the shortcomings of the previously mentioned models, aiming to enhance the precision of predictive analysis. A fusion model is created by training an LSTM model using the outputs of ARIMA, GARCH, and Random Forest models as features. ARIMA, a conventional statistical model, is ideal for predicting short-term outcomes, whereas machine learning models like LSTM are better suited for analyzing extensive datasets with nonlinear patterns. Therefore, the objective of this thesis is to investigate whether the results of the fusion of the two types of models outperform the separate models for their respective predictions. An LSTM model trained with the outputs of ARIMA, GARCH and Random Forest models as features will be used as the fusion model. To investigate the merits and demerits of these methods as well as the possibility of a hybrid approach to predicting stock prices, this study will evaluate each of these models individually and compare the prediction results of the individual models with their fusion model in order to explore the most accurate stock prediction model.

The major contributions of this paper are as follows:

1. Development of a novel fusion model combining ARIMA, GARCH, LSTM, and Random Forest to enhance stock price prediction accuracy.
2. An extensive evaluation and comparative analysis of traditional statistical models, machine learning models, and the proposed fusion model based on historical stock market data.
3. Demonstration of the effectiveness of feature engineering and integrated learning phases in improving prediction performance.
4. Provision of a more reliable and precise tool for stock market analysts and investors to make informed decisions in volatile financial markets.

The manuscript is structured as follows: Section 2 provides a review of related work in the fields of statistical modeling and machine learning methods for stock price prediction. Section 3 details the methodology, including the construction and development of the ARIMA, GARCH, LSTM, and Random Forest models, as well as the fusion model. Section 4 discusses the experimental procedure, data pre-processing, and evaluation metrics used in this study. Section 5 presents the results and comparisons of the models. Finally, Section 6 concludes the paper and suggests directions for future research.

2. RELATED WORK

- Statistical Modeling in Stock Price Prediction Model Study

ARIMA modeling is widely regarded as an exceptionally efficacious forecasting methodology within the domain of stock forecasting. As its predictions are derived from the values of the input variables and the error term, ARIMA forecasting does not necessitate the presupposition of any underlying model or associated equations. However, sophisticated nonlinear real-world problems may introduce some bias into the ARIMA model due to the fact that it is a linear regression model. However, it is generally observed that linear models outperform complex structural models when it comes to short-term forecasting (Ma, 2020). A method for forecasting the price of garlic was introduced by Yan W. et al. (Wang et al., 2022). This method utilized a combination of GARCH family models and LSTM. By constructing a GARCH family model, they acquired data on volatility characteristics, including volatility aggregation, of garlic price series. The LSTM network was employed to examine the complex nonlinear interactions between sequences of garlic prices and their intrinsic volatility, aiming to forecast subsequent garlic price trends. Resulting from the independence of the machine learning models, the fusion model proves to be effective. The study by Yan W. et al. and the anticipated stock price prediction model in this study share some similarities (Wang et al., 2022); this provides the inspiration for the concept of model fusion in this article.

- Research into machine learning methods for forecasting stock prices

LSTM is crafted as a variant of recurrent neural networks, especially skilled in handling and predicting major events in time series data marked by substantial intervals and periods. Jin, Z. et al. noticed the advantages of the LSTM model in analyzing the relationship between time series and adapted the LSTM model by using the attention mechanism to predict the closing price with greater precision (Jin et al., 2020). Park H et al. introduced a new stock prediction framework called LSTM-Forest, which combines LSTM and Random Forest to address the issue of overfitting in prediction models (Park, Kim and Kim, 2022). RF can handle very many features to avoid the overfitting problem and LSTM outperforms the decision tree in terms of temporal patterns, so their method can be used with ensure that no useful information is lost to mitigate overfitting and consequently enhance the efficacy of predictions. According to a study by Y. Ma et al., the majority of researchers who employed LSTM neural networks to forecast stock prices trained the networks using unprocessed stock data (Ma, Han and Fu, 2019). This approach resulted in the training model absorbing a substantial amount of noise and ultimately diminished the model's predictive performance. Principal Component Analysis (PCA) and Random Forest were employed to identify critical input features, thereby enhancing the stock price prediction performance. The outcomes demonstrate that the stock prediction model constructed using Random Forest and LSTM yields more accurate predictions.

3. METHODOLOGY

3.1. Model Construction

- ARIMA

In order to normalize non-stationary time series data, the ARIMA model integrates the concepts of differencing (I) and autoregressive (AR) and moving average (MA) models.Three primary parameters define the ARIMA model: p, d, q

p : the number of autoregressive terms. The autoregressive component is the effect of past values on current values in the model.

d : Represents the number of non-seasonal differencing operations applied to stabilize the time series.

q: Number of moving average terms. The moving average part models the effect of the forecast error term.

The model structure of ARIMA can be represented by the following equation:

$$(1-B)^d Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} + \epsilon_t \quad (1)$$

In this formulation, Y_t signifies the value of the time series at time t; B acts as the backward shift operator, $BY_t = Y_{t-1}$; $(1-B)^d$ represents the differencing operator to make the time series stationary; δ is a constant term; The autoregressive components are indicated by $\phi_1, \phi_2, \dots, \phi_p$, and the parameters of the moving average section are denoted as $\theta_1, \theta_2, \dots, \theta_q$; ϵ_t denotes the error at time t, which adheres to a normal distribution with a mean of zero and a variance of σ^2; e_t represents the forecast errors at time t.

- GARCH

Tim Bollerslev proposed the GARCH model in 1986. This model is a statistical framework used for analyzing time series data and is an extension of the ARCH model, which was initially developed by Robert F. Engle in 1982, which aimed to model the conditional variance fluctuations over time or their aggregation. The GARCH model, in essence, leverages historical fluctuations to predict autoregressive changes in volatility. It is effective in addressing issues like heteroskedasticity, volatility aggregation, leverage effect, asymmetric effects, and can closely capture the volatility dynamics of time series data (Wang et al., 2022).

A standard GARCH(p, q) model can be described using the following formula:

$$\begin{aligned} y_t &= \mu_t + \sigma_t \eta_t \\ \varepsilon_t &= \sigma_t \eta_t \end{aligned} \quad (2)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \quad (3)$$

The variables denoted as follows: η_t denotes the variables that are identically and independently distributed, $\eta_t \sim N(0,1)$; σ_t^2 outlines the conditional variance at time t; α_0 represents the constant; α_i represents the coefficients of the ARCH terms; β_i details the coefficients of the GARCH terms; and ε_t captures the residual at time t.

- LSTM

An enhanced iteration of the RNN method, LSTM was initially introduced by Hochreiter and Schmidhuber in 1997. LSTM introduces the mechanism of "gates", which can effectively control the forgetting and remembering of information, so that the network can still maintain a stable gradient flow in long sequences, and thus capture long-distance data dependencies.

A standard LSTM unit is composed of four key components: the forget gate, the input gate, the cell state, and the output gate (Park, Kim and Kim 2022 & Hochreiter et al., 1997). The architecture of a traditional LSTM is depicted in Figure 1. The activation value of the Forget Gate is f_t ; the activation value of the Input Gate is i_t , which determines the number of new candidate values added to the cell state from g_t; the activation value of the Output Gate is o_t; the cell state at the current time step is c_t, and the output of the current time step is h_t.

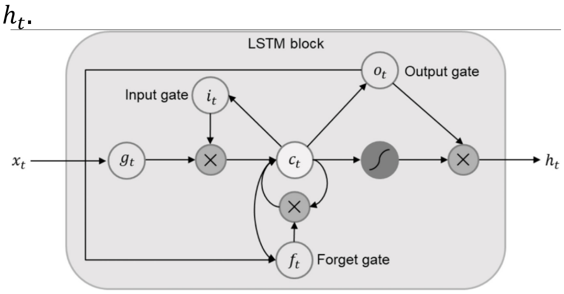


Figure 1. LSTM (Park, Kim and Kim, 2022).

- RF

The RF approach is extensively utilized for machine learning tasks in classification and regression. It is a technique called ensemble learning that improves the reliability and precision of a prediction model by combining the results of numerous decision trees to make a final conclusion. This unique framework enables the mitigation of flaws in individual classifiers and the amalgamation of the strengths of each classifier, hence enhancing prediction accuracy and managing overfitting.

The fundamental concept of random forest encompasses two key elements: the random selection of samples and the random selection of attributes from the data. Bootstrap Sampling is used to randomly choose data, whereas the random selection of features is achieved by determining which features are utilized as the training set during the division of each node in the decision tree.

The most important parameters for implementing the random forest model include:

Augmenting the quantity of decision trees boosts the stability of the model, but at the expense of increased computing requirements. The quantity of features considered during the splitting of each node in a decision tree influences the range of features sampled by the model, impacting both the bias and variance of the model's outcomes (Genuer et al., 2010 & Breiman, 2001).

3.2. Model Development

Machine Learning Models: LSTM models excel in capturing long-term correlations within the time series data of stock prices and adeptly managing nonlinear patterns. (Jin, Yang and Liu, 2020). On the other hand, RF models provide feature importance assessment to identify the key factors affecting stock prices and are able to avoid overfitting problems. The combination of these two models enhances the generalization ability of the model by effectively handling nonlinear patterns and high dimensional data. Therefore, the fusion model of Random Forest and LSTM may have better prediction results (Park, Kim and Kim 2022 & Ma, Han and Fu, 2019).

Time Series Models: ARIMA models are used to capture linear trends in time series, for example, to analyze and forecast specific seasonal characteristics and cyclical movements or trends of the stock market. Conversely, GARCH models are predominantly employed for the purpose of predicting the volatility of stock returns in the forthcoming periods. The convergence of GARCH and ARIMA models improves the accuracy of forecasts, specifically with regard to identifying patterns of volatility within the dataset.

3.2.1. Model Fusion Strategies

The GARCH model may determine the volatility characteristics of the series for the LSTM and ARIMA models. By combining these three models, more accurate prediction results can be obtained (Wang et al., 2022). The combination of the RF model and the LSTM model is capable of effectively handling nonlinear patterns and high-dimensional data, while also improving the model's generalization capacity (Bollerslev, 1986 & Engle, 1982).

3.2.2. Feature Engineering Stage

First, the historical stock price data are modeled and forecasted using ARIMA and GARCH models, respectively. The ARIMA model outputs the predicted future values and the GARCH model outputs the predicted volatility estimates.

Second, the historical stock price data are fitted with a random forest model in order to evaluate the significance of various features, including those produced by ARIMA and GARCH and other stock metrics.

3.2.3. Integration Learning Phase

First, integrating ARIMA and GARCH: The forecasts from the ARIMA and GARCH models are combined to form a comprehensive set of time series features.

Subsequently, the features identified by the random forest and the combined outputs from ARIMA and GARCH models are utilized as input data to enhance the LSTM model's training process.

Finally, this specifically tailored dataset is employed to train the LSTM model, which is then used to forecast stock prices.

4. EXPERIMENTAL PROCEDURE AND ANALYSIS OF RESULTS

4.1. Data and Data Pre-Processing

Data sets and experimental tools: The original data was collected from Yahoo Finance. The study used stock market data from Apple, Google and Amazon for a total of 2,571 days between January 2, 2014 and March 20, 2024. The data is in numerical format and includes detailed data for each day of the historical stock. All experiments were implemented in Python 3.11's Jupyter Notebook.

The acquired data will undergo a rigorous cleaning, normalization, and transformation process to make it suitable for time series analysis and machine learning modeling. Since the original stock data is very clean and complete, this study does not need to do much data preprocessing. Furthermore, to boost the precision of the forecasting model, this study incorporates supplementary indicators like Simple Moving Average (SMA), Exponential Moving Average (EMA), Bollingband, etc., to describe the dataset. Technical indicators offer vital insights into market patterns, volatility, and momentum, hence enhancing the precision and dependability of forecasts.

4.2. Evaluation Metrics

The study assessed the model's prediction ability by employing four distinct evaluation measures: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and R square. A prediction model's performance is considered better when the values of MAE, RMSE, and MAPE are lower. Additionally, a value of R square close to 1 indicates a high degree of fitting between the model and the data. The four evaluation indicators are calculated as follows.

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m} |(y_i - \hat{y}_i)| \tag{4}$$

$$\mathbf{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} (y_i - \hat{y}_i)^2} \tag{5}$$

$$\mathbf{MAPE} = \frac{100}{m}\sum_{i=1}^{m} |\frac{(y_i - \hat{y}_i)}{y_i}| \tag{6}$$

$$\mathbf{R^2} = \mathbf{1} - \frac{(\sum_{i=1}^{m} (y_i - \hat{y}_i)^2)/m}{\left(\sum_{i=1}^{m} \left(y_i - \bar{y}_i\right)^2\right)/m} \tag{7}$$

The objective of this research is to clarify the relative advantages and disadvantages of machine learning methods and time series models as they pertain to the forecasting of stock prices. Given its ability to capture both linear and non-linear patterns in stock price fluctuations, it is anticipated that the stacking model will outperform individual techniques. Through the development of a more precise and dependable instrument for forecasting stock prices, this study possesses the capacity to substantially advance the domain of financial analytics and ultimately aid investors in the process of making well-informed judgments.

4.3. Experimental Procedure

- ARIMA.

In this study, automatic ARIMA modeling was carried out through a auto_arima function, which autonomously searches for and chooses the most suitable model parameters. The function performs model comparisons based on the Akaike Information Criterion (AIC).The AIC aims to select parameter configurations that best fit the data while maintaining model simplicity. As shown in Table1 after a series of model configuration attempts, it was determined that the ARIMA (1,1,1) model was the optimal one.

Table 1. Configurations of ARIMA.

Argument	AIC
(1,1,1)	8414.499

The detailed statistical summary of the model is shown in Table 2, from which we know that the intercept term is insignificant, which means that the model does not find a statistically significant non-zero starting point or long-term trend in the data. In addition, both parameters ar.L1 and ma. L1 are significant. ar. The presence of L1 indicates a statistically significant link between the current value of the series and its prior value, ma.L1 is substantial, indicating a statistically significant link between the present value of the series and its stochastic shocks at the prior time point. However, the Ljung-Box test of the model indicates that the residuals do not exhibit autocorrelation, implying that the model captures most of the information in the time series.

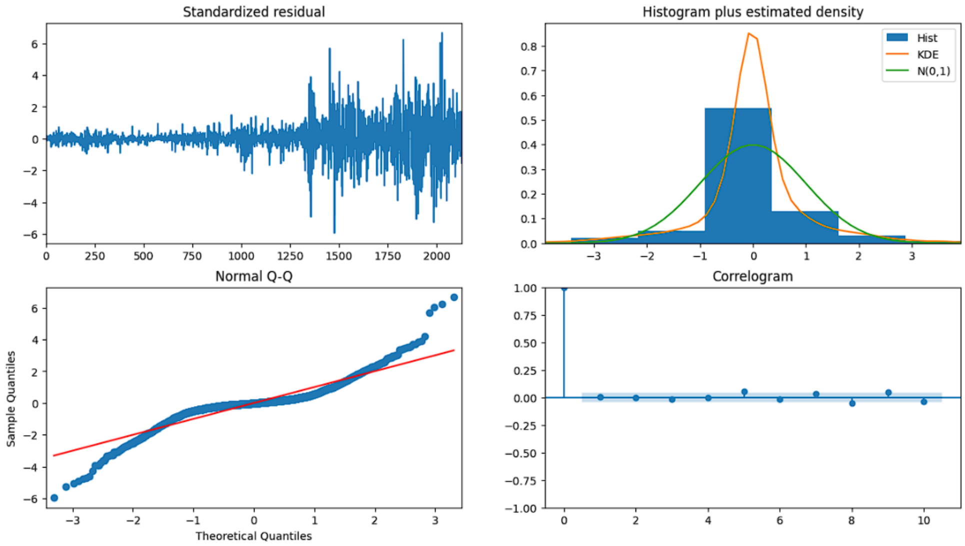
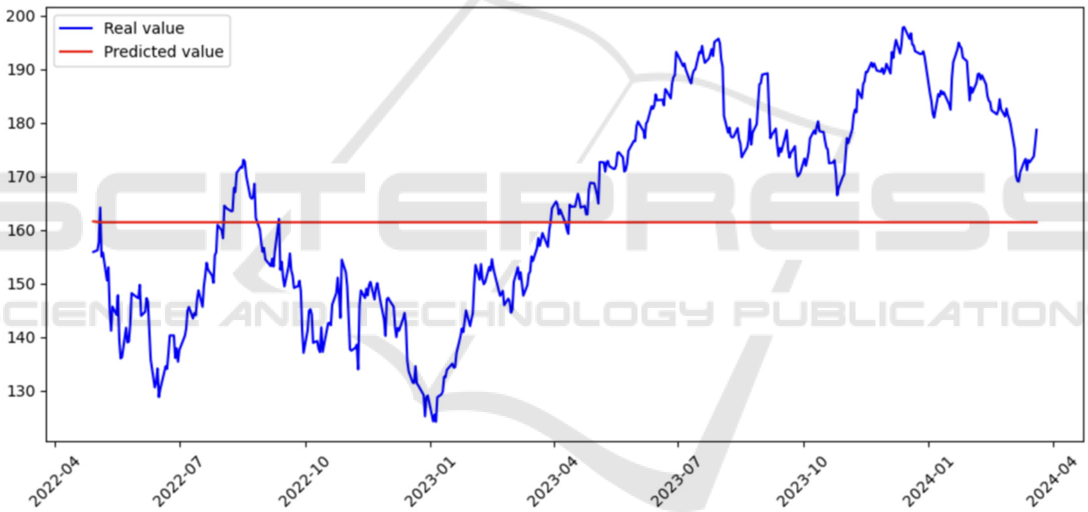


Figure 2. Result of autocorrelation (Picture credit: Original).



SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

Figure 3. the predicted stock prices using the ARIMA (Picture credit: Original).

Table 2. The detailed statistical summary of ARIMA.

Metrics	P-Value
intercept	0.118
ar.L1	0.014
ma.L1	0.004
sigma2	0.000
Ljung-Box(L1)	0.91
Jarque-Bera	0.00

Based on the results in table 2 and Figure 2 we know that the residuals of the model do not exhibit any significant autocorrelation. This outcome suggests the model effectively captures the structure of the data and that random fluctuations in the residuals are unpredictable. However, the model does not satisfy the assumption of residual normality and the non-normality of the residuals may constrain the model's forecasting ability, which implies that the predictive outcomes of the ARIMA method are unreliable, and we should be cautious about the predictive results of this model.

The outcome of the prediction set subsequent to training the ARIMA model with the training set is as shown in figure 3.

Figure 3 shows the predicted stock prices using the ARIMA model represented by the red curve, and the real stock prices by the curve line. The red cureve's lack of fit to the true value is evidently demonstrated to be a straight line; this indicates that the model is an inadequate predictor of long-term outcomes.

- GARCH:

This experiment constructs a number of GARCH models by employing a two-layer loop that varies the parameter combinations (p and q). The variables p and q in the GARCH models denote the lag order, which are used to determine the autoregressive and moving average terms in the model, respectively. In addition, we evaluated the performance of each model using AIC and selected the model parameters with the lowest AIC values.

After experimentation it was found that the lowest value of model AIC was found at p=1, q=1, so we selected this parameter for experimentation. To reflect the accuracy of the GARCH model, we compare the volatility of the original stock data with the predicted volatility, and the r-squared of the final model is 0.8756.

- LSTM:

In this study, we use two LSTMs and two Dropout layers to construct the model. Both LSTM layers have 50 units and the parameter return_sequences is set to True, which allows the LSTM to return so time-steps of consecutive outputs; the Dropout layer is set to 20% to reduce overfitting; and finally the model performs the prediction of the results through the fully-connected layer containing 25 neurons and an output layer.

The prediction set performs in figure 4.

Figure 4 depicts the date on the x-axis and the stock price on the y-axis. The red line depicts the predictions made by the LSTM model, and the actual values are denoted by the blue line. The graph indicates that while both lines follow a similar trend, they rarely align perfectly. Typically, the predicted values are slightly above the actual figures.

- Random Forest:

For this experiment, a total of 100 decision trees were established, with all other parameters being left at their default settings. Upon training the random forest model with the training set, the prediction performance of the prediction set is as shown in figure 5:

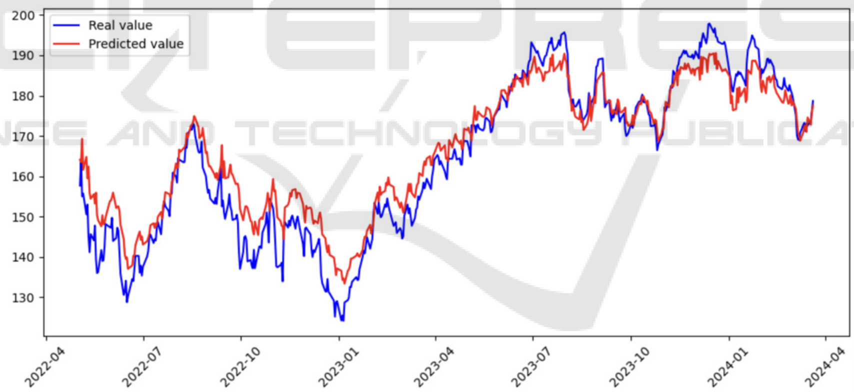


Figure 4. the predicted stock prices using the LSTM (Picture credit: Original).

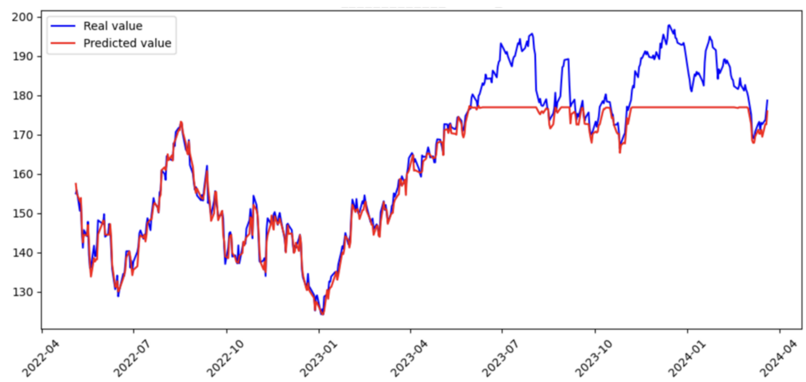


Figure 5. the predicted stock prices using the RF (Picture credit: Original).

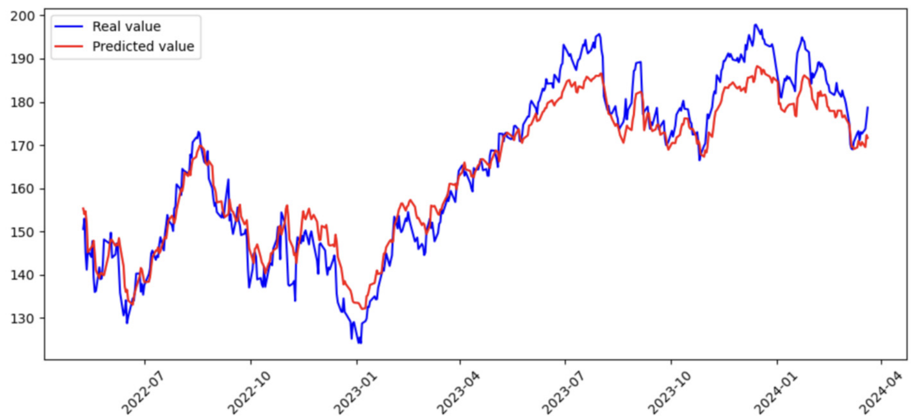


Figure 6. The predicted stock prices using the Fusion Model (Picture credit: Original).

Figure 5 demonstrates that the majority of the model's predictions largely coincide with the real values, indicating the model's superior capability to capture variations in stock prices. Nevertheless, at certain periods, significant disparities arise between the model's forecasts and the real values. These discrepancies could be attributed to market uncertainties or other factors that the model fails to account for. Consequently, it indicates that the RF model currently lacks the capability to accurately predict abrupt fluctuations in stock prices.

- Fusion Model:

To assess feature importance, we utilize the Random Forest method, incorporating the ten most crucial features along with the GARCH model's predictions into the previous LSTM model. The consequence obtained from the fusion model are illustrated in Figure 6.

- Model Comparisons:

Table 3 Results comparison.

	MAE	RMSE	MAPE	R^2
ARIMA	17.4000	19.6456	Nan	-0.0231
GARCH	0.0021	0.0027	14.6341	0.8756
LSTM	0.0169	10.4901	5.8656	0.7093
RF	4.3642	6.9100	2.4215	0.8734
Fusion Model	4.2696	5.1780	2.6105	0.9297

Table3 shows the fusion model surpasses the other models in accurately predicting the closing price of the stock. Its MAE, RMSE, MAPE are all the smallest, and the r-square is the closest to 1, which signifies that, relative to other models, fusion model has the minimum prediction error, and the predicted value is the closest to the original stock closing price, which is the best fitting effect.

5. CONCLUSION

This study examines the effective integration of time series and machine learning techniques in forecasting stock prices. It involves the development and analysis of ARIMA, GARCH, LSTM, Random Forest, and integrated models. It is demonstrated that while individual models possess their own strengths in specific contexts, hybrid models exhibit superior accuracy and reliability in stock price forecasting. The fusion model put forth in this study not only consolidates the benefits of each individual model but also enhances prediction accuracy significantly through feature engineering design and integrated learning phases. Specifically, Random Forest does exceptionally well in detecting important features, which enables the LSTM model to effectively capture extended relationships in time series data. Experimental findings showcase the superiority of the fusion model over single models both theoretically and in practical applications. Not only does it enhance forecasting accuracy, but it also bolsters the model's adaptability to sudden market fluctuations. This fusion approach equips stock market analysts and investors with a more precise and dependable decision-making tool, aiding them in making well-informed investment choices within the intricate and volatile financial markets.

It is important to acknowledge that while the fusion model exhibits remarkable forecasting capabilities for stock prices, its construction and training necessitate substantial computational resources and meticulous parameter tuning. Given the volatile and intricate characteristics of the stock market, it is evident that predictive models are incapable of entirely mirroring future pricing. This underscores the importance for investors to employ prudence when placing their trust in such models. Future research endeavors could focus on delving deeper into optimizing algorithm efficiency and refining parameter optimization models to further augment the precision and utility of stock price forecasting.

REFERENCES

Lu W, Li J, Wang J, et al. A CNN-BiLSTM-AM method for stock price prediction. Neural Computing Application, 2021, 33(10): 4741-4753.

Li J, Pan S, Huang L. A machine learning based method for customer behavior prediction. Tehnički vjesnik, 2019, 26(6): 1670-1676.

Sim H S, Kim H I, Ahn J J. Is deep learning for image recognition applicable to stock market prediction?. Complexity, 2019, 2019.

Ariyo A A, Adewumi A O, Ayo C K. Stock price prediction using the ARIMA model, 16th International Conference on Image Processing, Computer Vision and Machine Learning. IEEE, 2014: 106-112.

Sunny M A I, Maswood M M S, Alharbi A G. Deep learning-based stock price prediction using LSTM and bi-directional LSTM model, 2020 2nd NILES. IEEE, 2020: 87-92.

Nelson D M Q, Pereira A C M, De Oliveira R A. Stock market's price movement prediction with LSTM neural networks,2017 Joint Conference on Neural Networks. IEEE, 2017: 1419-1426.

Ma Y, Han R, Fu X. Stock prediction based on random forest and LSTM neural network,2019 19th International Conference on Control, Automation and Systems. 2019: 126-130.

Ma Q. Comparison of ARIMA, ANN and LSTM for stock price prediction, E3S WOC. EDP Sciences, 2020, 218: 01026.

Wang Y, Liu P, Zhu K, et al. A Garlic-Price-Prediction Approach Based on Combined LSTM and GARCH-Family Model. Applied Sciences, 2022, 12(22): 11366.

Jin, Z., Yang, Y. & Liu, Y. Stock closing price prediction based on sentiment analysis and LSTM. Neural Comput & Applic 32, 9713 – 9729 (2020).

Park H J, Kim Y, Kim H Y. Stock market forecasting using a multi-task approach integrating long short-term memory and the random forest framework. Applied Soft Computing, 2022, 114: 108106.

Ma Y, Han R, Fu X. Stock prediction based on random forest and LSTM neural network,2019 19th International Conference on Control, Automation and Systems. IEEE, 2019: 126-130.

Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

Genuer R, Poggi J M, Tuleau-Malot C. Variable selection using random forests. Pattern recognition letters, 2010, 31(14): 2225-2236.

Breiman L. Random forests. Machine learning, 2001, 45: 5-32.

Bollerslev T. Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 1986, 31(3): 307-327.

Engle R F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: JOTS, 1982: 987-1007.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS