Salud, Ciencia y Tecnología - Serie de Conferencias. 2025; 5:1320
doi: 10.56294/sctconf20251320

ORIGINAL

Application of ARIMA model and deep learning in forecasting stock price in Vietnam

Aplicación del modelo ARIMA y el aprendizaje profundo para pronosticar el precio de las acciones en Vietnam

Nguyen Thi Thanh Loan[1], Dang Ngoc Hung[1], Vu Thi Thuy Van[2]

[1]Hanoi University of Industry, Vietnam.
[2]National Economics University.

Cite as: Thanh Loan NT, Ngoc Hung D, Thuy Van VT. Application of ARIMA model and deep learning in forecasting stock price in Vietnam. Salud, Ciencia y Tecnología - Serie de Conferencias . 2025; 4:1320. https://doi.org/10.56294/sctconf20251320

Submitted: 13-05-2024 **Revised:** 14-08-2024 **Accepted:** 18-11-2024 **Published:** 01-01-2025

Editor: Prof. Dr. William Castillo-González

Corresponding author: Dang Ngoc Hung

ABSTRACT

Introduction: time series forecasting is essential in production, business, and policy-making. In Vietnam, statistical models have been used to forecast time series, such as foreign investment and consumer price index. However, only some studies have used deep learning models in predicting economic variables. The study aims to application of ARIMA model and deep learning in forecasting stock price in Vietnam
Method: this article aims to review and evaluate the predictive ability of the ARIMA and deep learning models when forecasting the prices of the five stocks with the largest capitalization from January 2018 to April 2023.
Results: the deep learning model's prediction results show that the deviation rate between the actual and predicted values is 3,3 %.
Conclusions: Businesses and stakeholders should increase the use of technology and artificial intelligence for forecasting to support decision-making

Keywords: Deep Learning; Stock Price; Vietnam; LSTM.

RESUMEN

Introducción: el pronóstico de series temporales es esencial en producción, negocios y formulación de políticas. En Vietnam, los modelos estadísticos se han utilizado para pronosticar series de tiempo, como la inversión extranjera y el índice de precios al consumidor. Sin embargo, solo algunos estudios han utilizado modelos de aprendizaje profundo para predecir las variables económicas. El estudio tiene como objetivo la aplicación del modelo ARIMA y el aprendizaje profundo en el pronóstico del precio de las acciones en Vietnam
Método: este artículo tiene como objetivo revisar y evaluar la capacidad predictiva de los modelos ARIMA y de aprendizaje profundo al pronosticar los precios de las cinco acciones con la mayor capitalización desde enero de 2018 hasta abril de 2023.
Resultados: los resultados de predicción del modelo de aprendizaje profundo muestran que la tasa de desviación entre los valores reales y predichos es 3,3 %.
Conclusiones: las empresas y las partes interesadas deberían aumentar el uso de la tecnología y la inteligencia artificial para el pronóstico para apoyar la toma de decisiones.

Palabras clave: Aprendizaje Profundo; Precio De Las Acciones; Vietnam; LSTM.

INTRODUCTION

Stock price prediction is a challenging problem due to the complex and volatile nature of stock prices. Both researchers and investors have shown interest in this area. Various methods, from traditional analytical techniques to artificial intelligence methods such as fuzzy logic and artificial neural networks, have been used to predict financial markets.[1,2]

Time series forecasting involves using statistical and modeling methods to predict future trends. Traditional econometric models can be effective but have drawbacks, including compromised performance due to lack of specificity, inability to identify complex fluctuations in data, and inaccurate long-term forecasts. One of them is the autoregressive integrated moving average (ARIMA) model. From the first day, this model has been widely used in many fields, such as statistics, estimation, and forecasting, because it is found that the model has many advantages.[3,4] High accuracy when forecasting while eliminating multicollinearity in the model is the most significant advantage of ARIMA. However, this model also has some disadvantages, which are: difficulty in establishing a standard model from a group of potential models; forecasting methods are subjective; theoretical models and structural relationships do not differ from some simple predictive models (Thomas & Stekler, 1983); The data must follow the assumptions of the linear regression model. Overcoming the limitations of traditional machine learning models, deep learning (DL) models are a significant step forward. Many studies show that deep learning models are widely applied in the forecasting and achieving high-accuracy results. DL is a part of machine learning (ML). In recent years, DL techniques have outperformed traditional models. In particular, DL is successfully applied when solving time series forecasting problems because DL uses a multi-layered structure of algorithms called neural networks that make data preparation faster and more efficient.

In deep learning, pre-trained models mostly consist of neural networks. There are three types of neural networks: artificial neural networks (ANNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs). Long Short-Term Memory (LSTM) is an improved version of RNN that makes it easier to remember past data. The LSTM network builds the model using the backpropagation method. LSTM is well-suited for classifying, processing, and predicting time series with an unknown delay since each forecasting model has pros and cons for getting the most accurate time series forecasting results.

ANN is often used to describe the nonlinear relationship between input and output, but local extremes can easily trap it. Additionally, ANNs have many free parameters that often require a trial-and-error approach to select.[5] Some studies use support vector regression (SVR) (Bishop & Nasrabadi, 2006), which has been shown to provide accurate results in stock price prediction problems.[1] However, choosing the right technical analysis indicator for each stock is not straightforward since an indicator that works well for one stock may not work for another.[6] Given the pros and cons of each forecasting model, how can we obtain the most accurate time series forecasting results?

In Vietnam, numerous websites and securities apps supported by machine learning can predict market trends through automatic analysis of trading data. Machine learning also provides a capital allocation strategy for investors to optimize profits, including robust disbursement during reliable waves and limiting transactions during downtrends or sideways movements. Securities companies generally follow this trend but need to be more synchronous and thorough. Furthermore, research on applying deep learning to stock price forecasting is necessary.

The primary objective of this study is to assess the forecasting capabilities of ARIMA and deep learning models for listed stocks in Vietnam. This evaluation will lay the groundwork for implementing ML forecasting models in investment.

Theories

Models

Model ARIMA

The model uses time series data, considering the past values of a particular variable as good indicators of its future value. Specifically, Yt represents the variable's value at time point t with Yt = f(Yt-1, Yt-2, ..., Y0, t).

The analysis aims to identify relationships between observed Yt values to forecast future values, particularly for short-term forecasting purposes.

Autoregressive model p - AR(p): In the autoregressive model, the process depends on the sum of the weights of the past values and the random noise term.

$Yt = \varphi 1Yt\text{-}1 + \varphi 2Yt\text{-}2 + ...+\varphi pYt\text{-}p +\delta +\varepsilon t$

Moving average model q – MA(q): In the moving average model, the process is described purely by the weighted sum of the lagged stochastics:

$Yt = \mu +\varepsilon t -\theta 1\varepsilon\ t\text{-}1 - \theta 2\varepsilon\ t\text{-}2 -...-\theta\ q\varepsilon t\text{-}q$

Moving Average Combined Regression Model - ARMA(p,q):

$Yt = \varphi 1Yt\text{-}1 + \varphi 2Yt\text{-}2 + ...+\varphi pYt\text{-}p +\delta +\varepsilon t - \theta 1\varepsilon t\text{-}1 -\theta 2\varepsilon t\text{-}2 -...-\theta q\varepsilon t\text{-}q$

https://doi.org/10.56294/sctconf20251320

Consider the stationarity of the series of observations

The AR and MA components of the ARIMA model require stationary time series, but many time series are nonstationary. Stationarity means that the mean and variance of a stochastic process of Yt do not change over time. Additionally, the price covariance between two time periods depends only on the time lag distance between these periods and not on the actual time at which the covariance is calculated. There are three ways to detect the stationarity of a time series: based on the graph of the time series, the graph of the sample autocorrelation function, or the Dickey-Fuller test.

ARMA(p,d,q) model recognition involves determining appropriate values of p, d, and q. Here, d represents the order of difference of the time series being investigated, p is the order of autoregression, and q is the order of moving average. The Box-Jenkins method is commonly used for evaluating the ARIMA model fit. Moreover, a time series generally has three main components: trend, period, and season, and the ARIMA model requires the time series to be stationary. A seasonal ARIMA model (SARIMA) has been developed to include the seasonal factor in the model. However, when forecasting in the long term, SARIMA model cannot provide accurate results if the time series still contains the remaining two components (trend and period). To improve the prediction quality in the long term, one can find a smooth function through the neural network. The original time series can be transformed into this smooth function by removing the seasonal element from the time series (Fathi, 2019).

Model LSTM

The LSTM network overcomes the long-term dependency issues of traditional RNNs. LSTM, introduced by [7], has a chain structure of network nodes similar to RNNs, but with a more complex internal structure consisting of four layers of interaction. The unique feature of the LSTM network lies in the cell state C, where the long-term weights of the model are stored. The parameters of the cell state C, hidden state h, and input at time t xt are fed to the network node and after processing through sigmoid activation functions, tanh, and vector operations, the output is the cell state C and hidden state h at time t, which will be used for the nodet+1 (figure 1).

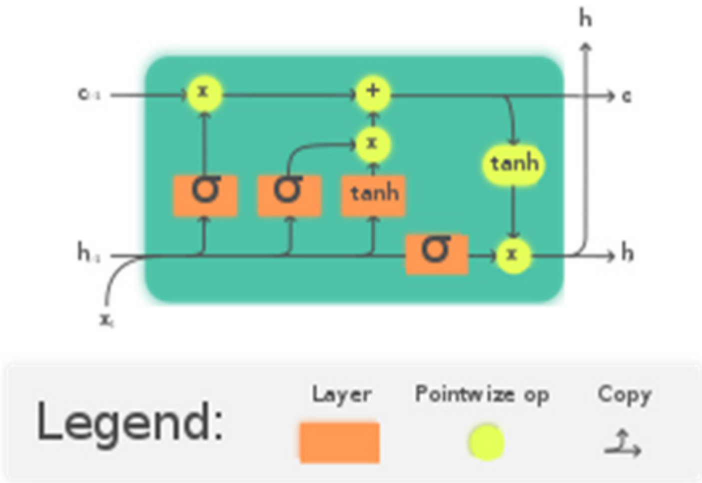

Figure 1. Structure of a node in an LSTM network

Some indicators of forecasting quality

Various indicators are utilized to evaluate the quality of a model. Some commonly used metrics for predicting model quality include:

Mean Absolute Error (MAE) measures the mean absolute error between predicted and actual values.

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|yi-\widehat{yi}|$$

We use MAE as a standard norm of order 1. The smaller the MAE, the better the model. However, MAE does not imply unit difference.

An estimator's Mean Squared Error (MSE) is the average of the square of the error. It measures the difference between the predicted and actual values. The MSE is a risk function corresponding to the expected value of squared error loss or quadratic loss. It is also the quadratic moment of the error.

$$MAE=\frac{1}{n}\sum_{i=1}^{n}(yi-\widehat{yi})^2$$

https://doi.org/10.56294/sctconf20251320

Root Mean Squared Error (RMSE) is the residual's standard deviation (prediction error). The residual measures distance from the regression line data points; The RMSE measures the spread of these residuals. In other words, it tells how concentrated the data is around the best-fit line. RMSE is commonly used in forecasting and regression analysis to verify experimental results. RMSE is a measure of how effective the model is. It does this by measuring the difference between predicted and actual values. The smaller the RMSE, the smaller the error, and the highest level of estimation that shows the model's reliability can be achieved.

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi-\hat{yi})}^{\,2}$$

The mean absolute percentage error (MAPE) is the mean fundamental percentage error of the forecasts. The better the forecast results, the smaller the MAPE. MAPE can be calculated as follows:

$$\text{MAPE}=\frac{1}{n}\sum_{i=1}^{n}\left|(Y_i-\hat{Y}_i)/Y_i\right|$$

Akaike Information Standards

One commonly used criterion for selecting a model is the Akaike Information Criteria (AIC) index. AIC is based on information theory and estimates the amount of information lost by a given model. AIC refers to a trade-off between model fit and simplicity, dealing with the risk of over- and under-fitting. Here, k is the number of parameters and is the maximum value of the full likelihood function of the model.

Literature reviews

Foreign studies have combined traditional machine learning models and deep learning for time series forecasting. According to [8] used this combination to forecast the price correlation coefficient of two separate stocks. The combined model ARIMA-LSTM has superior predictive power compared to traditional models.

Research conducted by [9] compared the predictive ability of five neural network models: backpropagation (BP), radial basis function (RBF), general regression neural network (GRNN), support vector machine regression (SVMR), and least squares support vector machine regression (LS-SVMR). The study used mean square error and mean absolute error as evaluation criteria and made stock price predictions for the Bank of China. Results showed that the neural network was consistently superior and more robust than the other four models.

Meanwhile [10] proposed DP-LSTM, a deep neural network, to predict stock prices. DP-LSTM is an LSTM-based neural network that reduces prediction errors and enhances strength levels. The proposed model achieves a 0,32 % improvement in mean MPA and an MSE improvement of up to 65,79 %. Doshi et al. (2020) used LSTM neural networks to forecast stock prices for four companies and designed a trading strategy that performed portfolio optimization. The LSTM model with a custom loss function performed better than a regression model like ARIMA.

According to [11], stock price prediction has been an essential area of study for a long time. While proponents of the efficient market hypothesis believe that it is impossible to predict stock prices, formal proposals demonstrate that correct modeling and the design of appropriate variables can lead to models using stock prices, and models of stock price movements can be predicted very accurately. Researchers have also worked on stock analysis techniques to identify patterns in stock price movements using advanced data mining techniques. The study used a combined modeling approach to predict stock prices by building machine and deep learning-based models. The results show that the most accurate model is the univariate model based on the LSTM using the previous week's data as input to predict the following week's NIFTY 50 time series open value.

In recent years, several studies have used machine learning algorithms and ANN in Vietnam's financial sector, such as [12,13,14]. However, few studies have used the LSTM model in stock price forecasting. Therefore, applying these tools to optimize and improve the predictability of stock quotes will have significant implications.

METHOD

Research data

The article utilized a dataset of 5 stocks with the highest market capitalization in Vietnam's stock market from January 2018 to April 2023, consisting of 1322 samples. The dataset is available on the Vietstock website. Table 1 presents an overview of the 5 stocks, which includes two stocks in the banking sector, one stock in the construction and real estate sector, and one in the energy sector.

Table 1. Research data

Company	Stock ticker	Field of activity	Market capitalization (trillion VND)	Time	Number of observations
Joint Stock Commercial Bank for Foreign Trade of Vietnam	VCB	Financial services	319,444	2/1/2018-20/04/2023	1322
Vinhomes Joint Stock Company	VHM	Real estate	214,670	2/1/2018-20/04/2023	1322
Vingroup company	VIC	Construction and Real Estate	220,445	2/1/2018-20/04/2023	1322
Joint Stock Commercial Bank for Investment and Development of Vietnam	BID	Finance and Insurance	164,907	2/1/2018-20/04/2023	1322
Vietnam Gas Corporation	GAS	Distribution of natural gas	210,917	2/1/2018-20/04/2023	1322

Figure 2 reflects the price trends of 5 stocks from Jan 2018 to Apr 2023. Each stock has different fluctuations. In Apr 2020, the stock price dropped due to Covid impact, then increased in early 2022, and decreased again in early 2023.

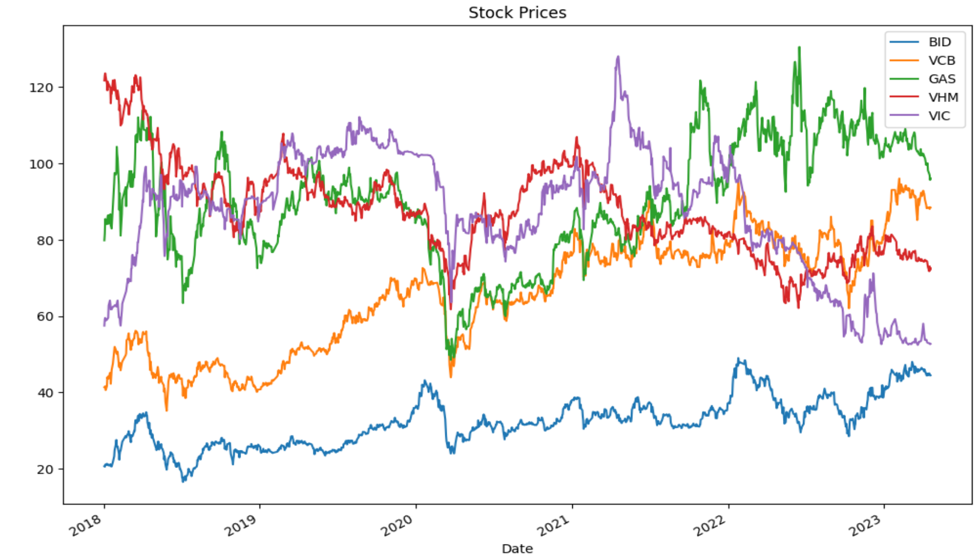

Figure 2. Price of 5 stocks from 01/2018 to 04/2023

Figure 3 displays the trading volume of 5 stocks from January 2018 to April 2023. The data indicates that the trading volume reached its peak in 2018, followed by a sharp decline from 2019 to 2021. The trading volume then experienced a sudden increase in 2022 before finally decreasing again in 2023.

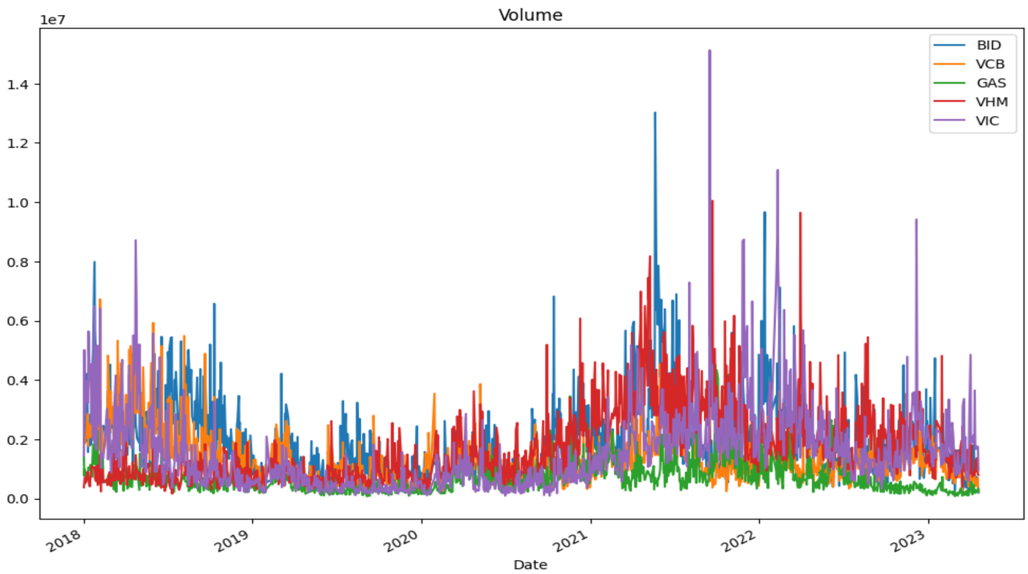

Figure 3. Trading volume of 5 stocks from 01/2018 to 04/2023

https://doi.org/10.56294/sctconf20251320

Research Methods

Forecasting method according to the ARIMA model

Start by using the Dickey-Fuller test to check if the time series is stationary. Since all of the P-values for the five stocks are more significant than 0,05, it can be concluded that the series is not static at the 5 % significance level. Next, perform a “grid search” using different parameter combinations. Each set of parameters will fit a SARIMA model and evaluate its quality. Finally, select the best ARIMA model for stock price forecasting.

Table 2. Test of stock price stoppage

Test		VCB	VHM	VIC	BID	GAS
Augmented Dickey-Fuller	Statistics ADF	-1,587	-2,899	-1,702	-2,207	-2,616
	Probability	0,490	0,045	0,430	0,20	0,09
Critical value	1 %	-3,435	-3,435	3,435	3,435	3,435
	5 %	-2,864	-2,864	2,864	2,864	2,864
	10 %	-2,568	-2,568	-2,568	-2,568	-2,568

LSTM method

After training a general model to fit the LSTM model dynamically, extensive hyperparameter tuning was performed. The hyper-parametric search included batch size, the maximum number of epochs, learning rate, hidden layer size, number of stacked LSTM layers, and dropout status. These parameters were tested by selectively scaling a single parameter while fixing the others. Related parameters (such as learning rate and the maximum number of epochs) were considered together. Compared to the simple method, LSTM worked very well, and dropout rates significantly impacted the model’s performance. Models performed better with less than four layers, and models with one layer had better results. The team tried different architectures that varied the number of layers and units in the layers until they got the best possible MAPE, and the model chosen is a sequential model of layers, 3 LSTM layers, 3 dropouts, and one dense layer.

```
Model: “sequential”
_________________________________________________________________
 Layer (type) Output Shape Param #
=================================================================
 lstm (LSTM) (None, 5, 200) 161600

 lstm_1 (LSTM) (None, 200) 320800

 dense (Dense) (None, 200) 40200

 dense_1 (Dense) (None, 100) 20100

 dense_2 (Dense) (None, 50) 5050

 dense_3 (Dense) (None, 5) 255

=================================================================
Total params: 548 005
Trainable params: 548 005
Non-trainable params: 0
_________________________________________________________________
```

Figure 4. Parameters of the LSTM

Maintaining the order of observations when dividing training and test data is essential. Choose the first 90 % of observations for training and testing, and reserve the remaining 10 % for testing instead of random selection. The dataset has 1322 observations split into two parts. Part 1, which includes the first 1184 observations, is used to build the LSTM model. Part 2, which consists of the remaining 138 observations, is used to verify the accuracy of the LSTM model.

To calculate, authors normalized data in interval [0;1] using the MinMaxScaler method from the Sklearn library. Then, they created an LSTM model with 50 neurons and 4 hidden layers using the MSE function and Adam optimal iterative algorithm 1000 times.

https://doi.org/10.56294/sctconf20251320

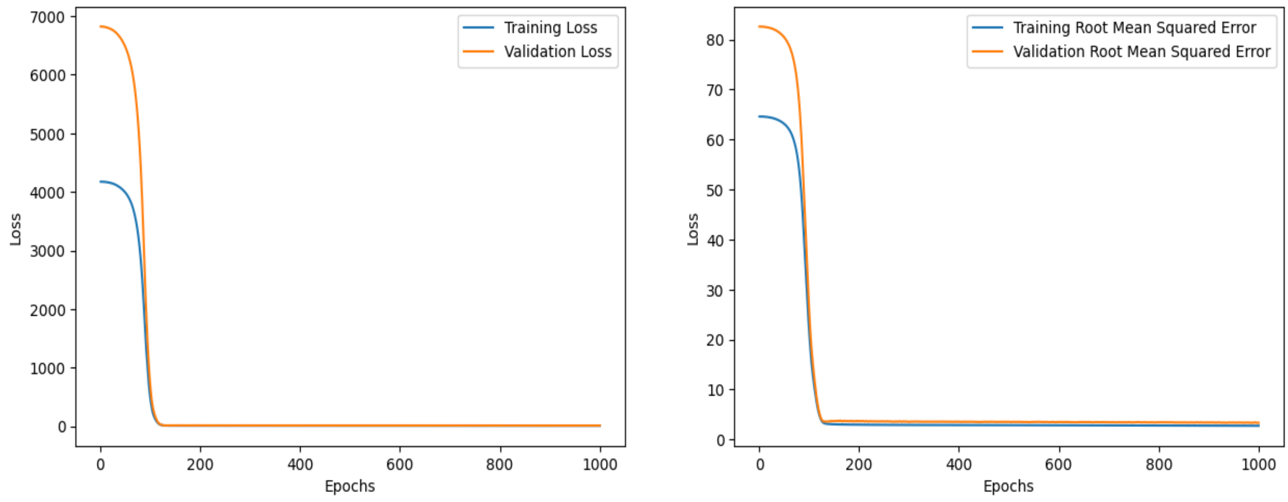

Figure 5. Evaluation of model quality

RESULTS AND DISCUSSION

After testing stationarity in Section 4 with the Dickey-Fuller test, we use a “grid search” to determine ARIMA/SARIMA model parameters that fit and rate their overall quality. Table 2 presents the best ARIMA/SARIMA model for each stock. Using the Root Mean Squared Error (RMSE) evaluation criteria, GAS stock has the lowest standard deviation of balance (5,59), and VCB has the highest (13,22). We continue to use the MAPE evaluation criteria when applying the ARIMA/SARIMA model to each stock. Results indicate that GAS stock has the lowest predicted versus actual value (4,74 %), and VCB stock has the highest (11,16 %). Thus, the ARIMA/SARIMAX model provides accurate forecast results ranging from 88,94 % to 95,26 %.

Table 3. Forecasting and evaluation results according to the ARIMA/SARIMA model

	Criteria	VCB	VHM	VIC	BID	GAS
1	ARIMA/ SARIMA	(0, 1, 0))	(0, 1, 0)	(3, 2, 3)	(0, 1, 0)	(0, 1, 0)
2	AIC	-5995,59	-6412,30	-6059,05	-5394,55	-5389,05
3	SIC	-5990,52	-6407,22	-6023,53	-5389,48	-5383,98
4	HQIC	-5993,68	6410,38	-6045,66	-5392,64	-5387,14
5	Loglikelihood	2998,80	3207,15	3036,53	2698,28	2695,53
6	RMSE	13,22	6,09	11,78	8,46	5,59
7	MAPE (%)	11,16	5,50	10,79	7,33	4,74

Section 1 analyzes and presents the efficiency of forecasting using the LSTM model to predict the price of 5 stocks. Figure 4 shows that the LSTM outperforms the ARIMA/SARIMA model in short-term forecasting, with lower RMSE/MAPE values. The LSTM model performs best when RMSE and MAPE values are lower than the ARIMA/SARIMA model. The combined forecast has a 4,6 % decrease in MAPE and a 6,2 decrease in RMSE compared to ARIMA.

Table 4. Comparison table of forecast results according to ARIMA and LSTM

Company	Ticker	Root Mean Squared Error (RMSE)			Mean absolute percentage error (MAPE)		
		ARIMA	LSTM	Compare LSTM vs ARIMA	ARIMA	LSTM	Compare LSTM vs ARIMA
Joint Stock Commercial Bank for Foreign Trade of Vietnam	VCB	13,22	3,35	-9,87	11,2 %	3,3 %	-7,9 %
Vinhomes Joint Stock Company	VHM	6,09	2,21	-3,88	5,5 %	2,2 %	-3,3 %
Vingroup company	VIC	11,78	3,39	-8,39	10,8 %	4,5 %	-6,3 %
Joint Stock Commercial Bank for Investment and Development of Vietnam	BID	8,46	2,19	-6,28	7,3 %	4,4 %	-2,9 %
Vietnam Gas Corporation	GAS	5,59	3,03	-2,56	4,7 %	2,1 %	-2,7 %
Medium		9,03	2,83	-6,20	7,9 %	3,3 %	-4,6 %

https://doi.org/10.56294/sctconf20251320

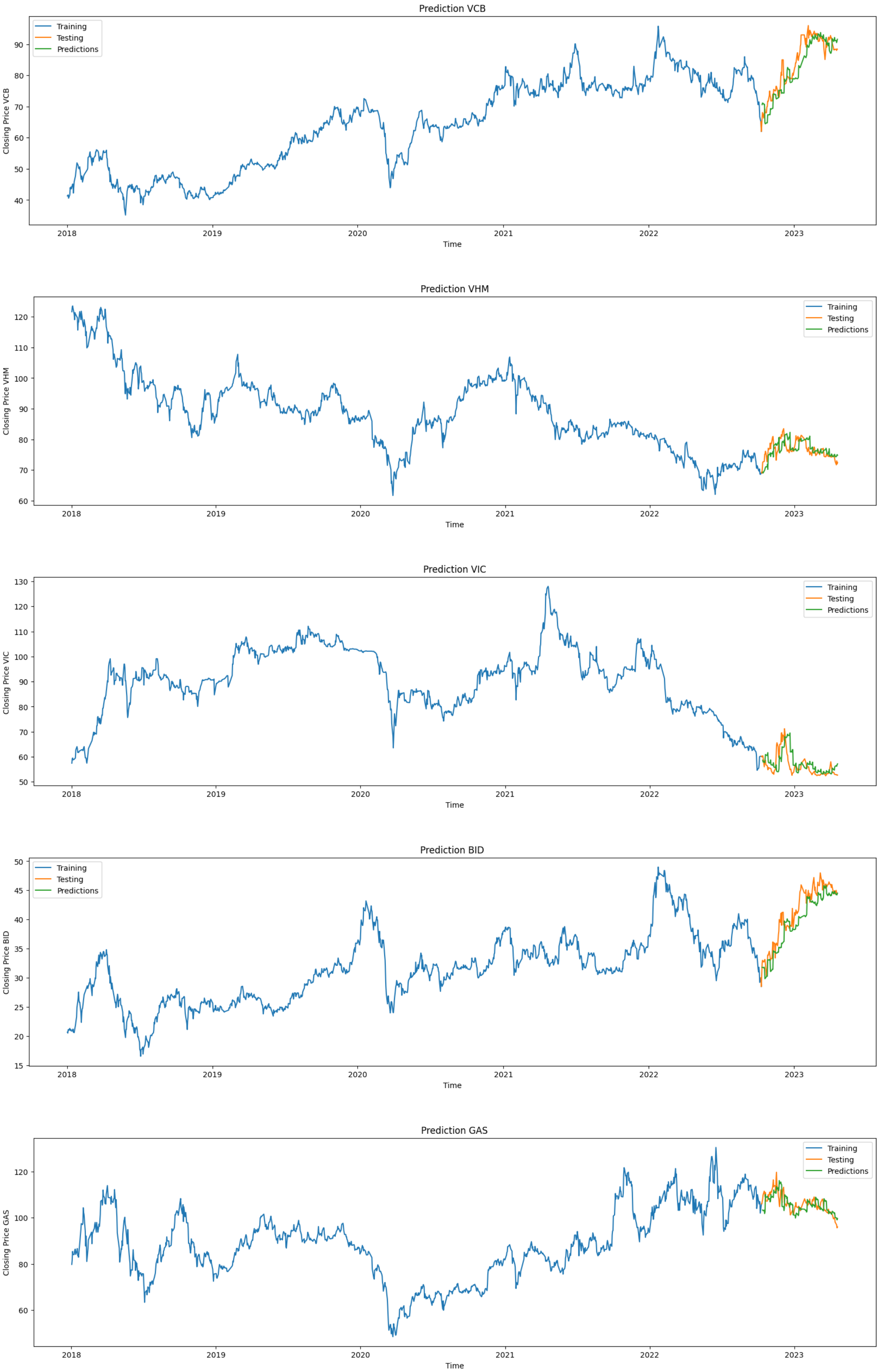

Figure 6. Forecast results for 5 stocks

https://doi.org/10.56294/sctconf20251320

In figure 6, the red line represents forecast values, the blue line represents actual values, and the orange line represents evaluated values. There is no significant difference between the actual and forecast values regarding trends. The forecast results for banking stocks (VCB, BID) tend to increase, while the stock prices of construction and real estate companies (VHM, VIC) and energy industry stocks (GAS) tend to decrease.

The two methods of ARIMA and deep learning bring great benefits in stock price forecasts. Arima is suitable for short -term forecasts and linear data, while deep study has a remarkable advantage in handling nonlinear relationships, large and multi -dimensional data. The combination of both methods can optimize the forecast results, bring higher accuracy and better adaptability in complex financial markets.

CONCLUSIONS AND RECOMMENDATIONS

Forecasting models based solely on past time series data can yield near-accurate stock price forecasts. However, stock prices are also influenced by various factors, such as those in the money and real estate markets and investor sentiment. Furthermore, the ARIMA model's deviation in forecast value is caused by its consideration of only past volatility time series. Additionally, the stock price chain in the Vietnamese market is assessed as needing to fully reflect information that can affect the price. The Vietnamese stock market's efficiency level is low, and its ability to reflect price information is still low, making forecasting errors inevitable, even when using different methods. While other forecasts like fundamental or technical analysis may be encountered, applying the deep learning method has produced accurate predictions. The average deviation between actual and forecast value is 3,3 %, depending on the stocks, but the variation does not exceed 5 %.

Investors must utilize reliable methods when forecasting stock prices to make informed decisions when buying or selling securities. The LSTM model has been proven to be highly accurate and dependable when predicting stock price movements. It is an essential tool for any investor looking to make intelligent and profitable investment decisions.

We must act to amplify the size of our data sets (e.g., stock markets, oil prices, gold prices, exchange rates) for further investigation or delve deeper into analyzing the SARIMA model, researching data seasonality, and exploring traditional machine learning models. We must discover a way to run the ARIMA model without sampling data due to computational limitations so that we can make a more precise comparison between ARIMA and LSTM. Additionally, we need to conduct further research to comprehend the trends established by LSTM patterns fully. To prove the effectiveness of our proposed method, we will be testing the LSTM on various Vietnamese stocks.

While ARIMA and machine learning both have their own advantages in forecasting stock prices, they both have limitations related to the volatility and complexity of the stock market. It is important to use these models in conjunction with other analyses (fundamental analysis, technical analysis, market sentiment analysis) to increase forecasting accuracy.

REFERENCES

1. Hsu S-H, Hsieh JP-A, Chih T-C, Hsu K-C. A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression. Expert Systems with Applications. 2009;36(4):7947-51.

2. Abraham A, Nath B, Mahanti PK, editors. Hybrid intelligent systems for stock market analysis. Computational Science-ICCS 2001: International Conference San Francisco, CA, USA, May 28—30, 2001 Proceedings, Part II 1; 2001: Springer.

3. Box GE, Pierce DA. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of the American statistical Association. 1970;65(332):1509-26.

4. Bianchi L, Jarrett J, Hanumara RC. Improving forecasting for telemarketing centers by ARIMA modeling with intervention. International Journal of Forecasting. 1998;14(4):497-504.

5. Sapankevych NI, Sankar R. Time series prediction using support vector machines: a survey. IEEE computational intelligence magazine. 2009;4(2):24-38.

6. Ince H, Trafalis TB. Kernel principal component analysis and support vector machines for stock price prediction. Iie Transactions. 2007;39(6):629-37.

7. Schmidhuber J, Hochreiter S. Long short-term memory. Neural Comput. 1997;9(8):1735-80.

8. Choi HK. Stock price correlation coefficient prediction with ARIMA-LSTM hybrid model. arXiv preprint arXiv:180801560. 2018.

https://doi.org/10.56294/sctconf20251320

9. Song Y-G, Zhou Y-L, Han R-J. Neural networks for stock price prediction. Journal of Difference Equations and Applications 2018:1-14.

10. Li X, Li Y, Yang H, Yang L, Liu X-Y. DP-LSTM: Differential privacy-inspired LSTM for stock prediction using financial news. arXiv preprint arXiv:191210806. 2019.

11. Mehtab S, Sen J, Dutta A, editors. Stock price prediction using machine learning and LSTM-based deep learning models. Machine Learning and Metaheuristics Algorithms, and Applications: Second Symposium, SoMMA 2020, Chennai, India, October 14–17, 2020, Revised Selected Papers 2; 2021: Springer.

12. Ngoc Hung D, Thuy Van VT, Archer L. Factors affecting the quality of financial statements from an audit point of view: A machine learning approach. Cogent Business & Management. 2023;10(1):1-25.

13. Hung DH, Binh VTT, Hung DN, Ha HTV, Ha NV, Van VTT. Financial reporting quality and its determinants: A machine learning approach. International Journal of Applied Economics, Finance and Accounting. 2023;16(1):1-9.

14. Dang NH, Van Vu TT, Le Dao TN. Accounting information and stock returns in Vietnam securities market: Machine learning approach. Contabilidad y Negocios. 2022;17(33):94-118.

FINANCING
No financing.

CONFLICT OF INTEREST
None.

AUTHORSHIP CONTRIBUTION
Conceptualization: Nguyen Thi Thanh Loan, Dang Ngoc Hung.
Data curation: Dang Ngoc Hung.
Formal analysis: Nguyen Thi Thanh Loan.
Research: Vu Thi Thuy Van.
Methodology: Nguyen Thi Thanh Loan, Dang Ngoc Hung.
Project management: Nguyen Thi Thanh Loan.
Resources: Dang Ngoc Hung, Vu Thi Thuy Van.
Software: Dang Ngoc Hung.
Supervision: Dang Ngoc Hung, Vu Thi Thuy Van.
Validation: Nguyen Thi Thanh Loan.
Display: Dang Ngoc Hung, Vu Thi Thuy Van.
Drafting - original draft: Nguyen Thi Thanh Loan, Dang Ngoc Hung, Vu Thi Thuy Van.