NATIONAL ECONOMICS UNIVERSITY

FINAL REPORT

DSEB STUDENTS ESSAY PROJECT 2023

TOPIC

Forecasting the VN30 index: Comparing Prediction Performance Using ARIMA and LSTM

Ha Noi, 11/2023

NATIONAL ECONOMICS UNIVERSITY

FINAL REPORT

DSEB STUDENTS ESSAY PROJECT 2023

TOPIC

Forecasting the VN30 index: Comparing Prediction Performance Using ARIMA and LSTM

Team member	**Contribution**
Duc An Vu (Team leader)	: %
Quoc Minh Le	: %
Bang Anh Nguyen Ngoc	: %
Quoc Huy Nguyen	: 15%

Instructors: Tuan Long Nguyen

Ha Noi, 11/2023

Forecasting the VN30 index: Comparing Prediction Performance Using ARIMA and LSTM

Duc An Vu, Quoc Minh Le, Quoc Huy Nguyen, Bang Anh Nguyen Ngoc

National Economics University, Hanoi, Vietnam

Abstract

The stock market has increasingly become a prevalent investment avenue, offering significant potential for investors. Accurately forecasting market trends is crucial for investor decision-making. This research aims to evaluate the performance of Time series and Deep learning techniques specifically, the SARIMA and LSTM models in forecasting the VN30 index, which is indicative of the top 30 Bluechip stocks on the HoSE, ranked by market capitalization and liquidity. The data set includes the VN30 index from January 2009 to September 2023, downloaded from Investing.com. The results indicate that the LSTM model surpassed the SARIMA model, with a Mean Absolute Percentage Error (MAPE) of approximately 1%, effectively capturing the fluctuation of the index and offering valuable market trajectory insights for investor decision-making.

Keywords: LSTM, SARIMA, Forecasting, VN30 Index

1. INTRODUCTION

Since the early 2000s, the Vietnamese stock market has evolved into a significant investment medium, attracting both institutional and individual investors. As the market has developed, the potential for profit has increased alongside a corresponding rise in risk. The majority of investors tend to make stock transactions based on sentiment and are often influenced by short-term information, leading to erratic stock price movements. Such volatility impacts the stability and sustainable development of Vietnam's stock market. [1]

After a prosperous year in 2021, the stock market experienced a significant correction in 2022. On 11/15/2022, the VN30 Index recorded a historic low at closing, reaching 904 points, which represents a decline of over 42% from its peak of 1572 points achieved on 11/25/2021 (based on actual data collected). The market showed signs of a mild recovery in 2023, yet it remained relatively subdued. With such fluctuations, predicting the general market trend is of paramount importance for investor decision-making.

Globally, numerous studies have utilized the SARIMA (Seasonal AutoRegressive Integrated Moving Average) model to forecast time series data, providing deep insights into market fluctuation patterns. In Vietnam, research on the SARIMA model for forecasting the VN-Index is still relatively scarce, and there is even less focus on leading stock groups like the VN30 or HNX30 indices. Therefore, we aim to delve deeper into this matter, exploring the application of the SARIMA model in the context of Vietnam's stock market.

However, to effectively capture the index's volatility, we have found that relying solely on the SARIMA model is insufficient. Therefore, our team has decided to employ the LSTM Deep Learning model, which has been proven to yield accurate predictions in various studies involving time series data, such as cryptocurrency [2], Consumer Price Index (CPI) [3], and Gross Domestic Product (GDP)[4].

Our objective is not only to apply these methods to the data from the Vietnamese stock market but also to compare the forecasting efficacy between the two models. This endeavor will not only clarify the capability of each model in capturing and analyzing market trends but also aid investors in making more precise investment decisions based on robust data analytics and advanced technical analysis.

For this study, monthly data frequencies were selected, considering the operational schedule of the Vietnamese stock market, which is active on weekdays, from Monday to Friday. The adoption of daily data could introduce complexities in data processing, and the focus of this research is not on granular day-to-day market movements. Employing monthly data thus ensures a more stable estimation framework while adequately addressing the objectives of this research.

2. METHODOLOGY

The primary objective of our study is to forecast the dynamics of the Vietnamese stock market by predicting the VN30-Index. This is achieved by analyzing data sourced from Investing.com. The dataset encompasses 3,678 entries across seven categories: Date, Open, High, Low, Close, Adjusted Close, and Volume, covering the period from January 2009 to September 2023. In the initial phase of data preparation, we conducted a preprocessing step. This involved transforming the dataset to a monthly frequency to circumvent issues related to missing values, acknowledging the Vietnamese stock market's operational pattern, which is active only on weekdays. This approach of using monthly data mitigates complexities in data handling while still aligning with our research objective to forecast the dynamics of the Vietnamese stock market, without delving into the intricacies of daily market fluctuations.

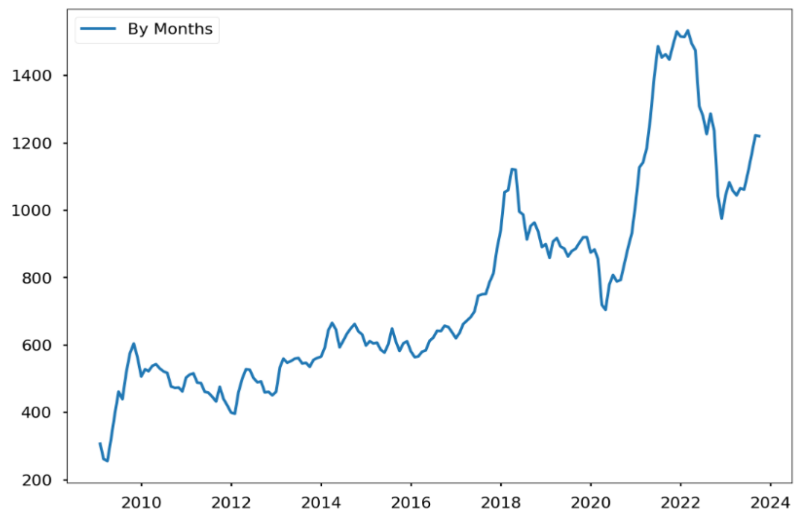

Figure 1. VN30-Index by month

Subsequently, the dataset was subjected to a thorough visual examination to discern underlying patterns and trends. Our analysis employed two distinct forecasting methods: SARIMA and LSTM, focusing on the VN30's closing index. Figure 1 illustrates the graphical representation of the VN30 closing index data, providing an initial overview for the subsequent analyses. The effectiveness of the SARIMA and LSTM models is assessed through two statistical metrics: Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). The computation of these metrics is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}$$

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

where:

$$n\text{: } amount\ of\ data$$

$$\widehat{Y}_i\text{: } Predicted\ value$$

$$Y_i\text{: } Actual\ value$$

2.1. ARIMA

The ARIMA (Autoregressive Integrated Moving Average) model is a widely-used forecasting tool for time series data. It gained popularity after being developed by Box and Jenkins in 1970, and is often referred to as the Box-Jenkins method. Its forecasting results depend on a chain of past values. The ARIMA method quantitatively analyzes the correlation between observed data to provide an informed forecast model. This involves the stages of model identification, parameter estimation from observed and test data, and using estimated parameters to find the appropriate model.The ARIMA method is applied to stationary data or stationary time series. Stationary data refers to data that fluctuates around a consistent long-term mean, maintaining a constant variance over time[5],[6].

This data exhibits self-correlation, where autocorrelation coefficients gradually decrease as the lag increases. In a stationary sequence, both the mean and variance remain constant over time. ARIMA is comprised of three main components, ARIMA(p, q, d):

- AR (p): Autoregressive
- I (q): Integrated (difference sequence)
- MA (d): Moving Average

To apply the ARIMA method, five steps are followed: Autoregressive, Integrated, and Moving Average. The AR part is used to model the correlation between observations at different points in time. The I part is used to make the time series stationary, and the MA part is used to model the dependency between an observation and a residual error from a moving average model applied to lagged observations. These steps help in identifying patterns and trends in the time series data and create a helpful time series model for future forecasting.[6]

2.1.1. Autoregressive (ARp)

In a study conducted by J. Sun in 2019[5], it was noted that in the AR (Autoregressive) model of order p, the future value of a variable is predicted through a linear combination of its past values, a constant, and a random error term. The equation (1) represents this AR model.

$$y_t = C + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \ldots + \Phi_p y_{t-p} + \varepsilon_t \qquad (1)$$

2.1.2. Integrated (I)

The behavior of a time series can be influenced by the accumulated impact of various processes. For instance, the condition of stock levels continuously changes due to consumption and supply dynamics. However, the average stock levels over time largely depend on the aggregate effect of these moment-to-moment changes occurring between inventory assessments. Even though short-term stock figures might experience significant variations around this average, the overall trend of the series remains stable over a longer period. This kind of time series, shaped by the cumulative impact of an ongoing activity, is categorized as an integrated process.[7]

Despite potential erratic movements in a series, the variations between consecutive observations might be minimal or might fluctuate around a steady mean when viewed at different intervals. This characteristic, where the series of differences demonstrates consistency, is essential from the standpoint of statistical analysis of time series. Integrated processes represent a fundamental example of non-stationary series.An integrated process is defined by equation (2):

$$y_t = y_{t-1} + \varepsilon_t \qquad (2)$$

where the random perturbation εt is a white noise.

2.1.3. Moving Average (MAq)

Similar to the AR (p) model, which forecasts by considering the values of previous data points in the series, the MA (q) model, as shown in mathematical equation (3), utilizes past errors as explanatory variables for prediction.

$$y_t = C + \varepsilon_t - \Phi_2 \varepsilon_{t-2} - \ldots - \Phi_q \varepsilon_{t-q} \qquad (3)$$

C is a constant or average, is a parameter of the model, and q is an order moving average.

2.1.4. Non Seasonal ARIMA

The basic model in the time series analysis is the ARIMA model. It is a combination of two processes – autoregressive (AR) and moving average (MA), which are weighted delayed random components[8]. When dealing with a stationary series, the ARIMA model can be formulated as follows:

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j} \qquad (4)$$

2.1.5. Model Seasonal Autoregressive Moving Integrated Average SARIMA(p,d,q)(P,D,Q)s

The SARIMA forecasting model is developed from the ARIMA model when the data series exhibits seasonality[9]. This is a phenomenon where the time series shows oscillation and repetition through each year[10]. The seasonal lags of yt are included in the ARIMA model to become the SARIMA model. At this point, the general SARIMA model is in the form of SARIMA(p,d,q)(P,D,Q)s, where P and Q are respectively the order of the seasonal AR and MA components, D is the order of seasonal differencing, and s is the number of periods in a cycle (in this study we use s = 12 for monthly data series).

Testing ACF and PACF at lags that are multiples of the seasonal period, s, is also crucial in determining appropriate P and Q values for the model. For the seasonal Moving Average (MA) component, the ACF chart typically shows distinct peaks at seasonal lags, while the PACF chart will display similar peaks for the seasonal AutoRegressive (AR) component. The values of d and D, the standard and seasonal differencing orders, are determined by the number of times differencing is needed to make the data series stationary.

2.1.6. Autocorrelation and Partial Autocorrelation Function (ACF and PACF)

In this study, ACF and PACF were analyzed to identify the suitable model for the VN30 time series data. These statistical measures reflect the relationships among observations within the time series. The ACF and PACF plots are created by plotting the correlation coefficients at successive lags.

The ACF plot displays the observed correlations at different lag values, with the x-axis representing the lags and the y-axis showing the correlation coefficients, ranging from -1 for negative correlation to 1 for positive correlation[6].

The PACF chart (Partial Autocorrelation Function) provides an overview of the correlation by excluding the influence of previous lag values. In other words, this chart analyzes the correlation at a certain lag without considering the influences from shorter lags.

To determine the optimal model, ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) were analyzed through Augmented-Dickey-Fuller statistical tests. The model selection process is based on evaluating various criteria, including the impact of model parameters, error indices, and information criteria such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan-Quinn Information Criterion (HQIC). These criteria help to determine the appropriateness and complexity of the model, aiming to find a balance between accuracy and simplicity of the model[6].

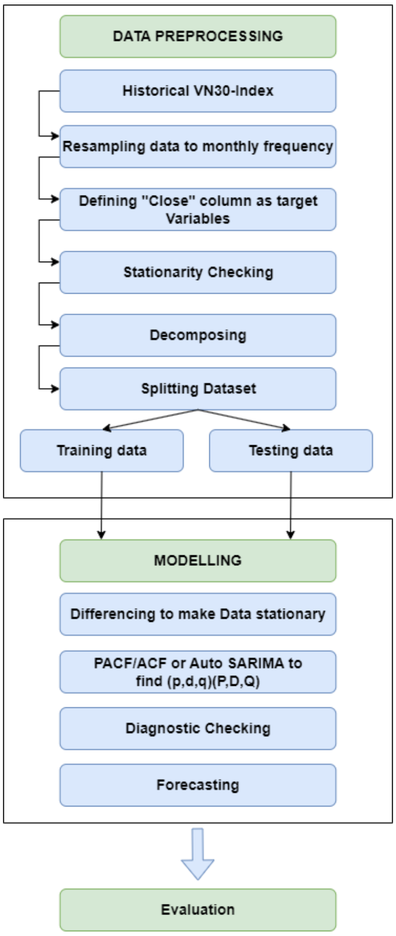

Figure 2. SARIMA Framework structure

For the SARIMA methodology, we adopted a framework structure comprising data preprocessing, model formulation, and performance evaluation, as illustrated in Figure 2.

2.2. LSTM

Long Short-Term Memory (LSTM) is a subclass of recurrent neural network (RNN) developed to analyze and predict time series data utilizing historical information[11]. In contrast to conventional RNNs, where the gradient tends to vanish due to the multiplication of small weights across multiple steps, LSTMs consist of memory blocks known as cells. These cells are linked through layers, enabling better information retention and processing.

The cell state in an LSTM, represented as Ct, and the hidden state, denoted as Ht, hold the cell's information. This data is regulated by gates using activation functions like sigmoid and Tanh. The sigmoid function yields values between 0 and 1, with 0 indicating complete blockage and 1 allowing full passage of information. These gates typically use the previous time step's hidden state (ht−1) and the current input (Xt). They execute a full multiplication of the input with a weight matrix (W) and add bias to enhance the output. LSTM includes three primary gates: the input gate (it), the forgetting gate (ft), and the output gate (ot). The input gate decides whether to allow new information in, the forgetting gate removes irrelevant data, and the output gate determines the final output. The functioning of these gates is defined by specific equations[11] as follows:

$$i_t = \sigma_g(w_i x_t + v_i h_{t-1} + b_i)$$

$$f_t = \sigma_g(w_f x_t + v_f h_{t-1} + b_f)$$

$$o_t = \sigma_c(w_c x_t + v_c h_{t-1} + b_c)$$

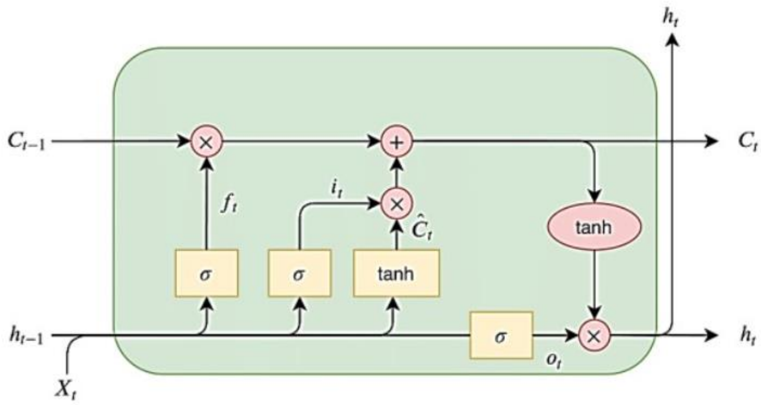

Figure 3. LSTM Cell

In each LSTM module, the inputs include Xt (the current input), Ht-1 (the hidden state from the previous step), and Ct-1 (the cell state from the previous step). The outputs are Ht (the current hidden state) and Ct (the current cell state). The input gate selectively allows specific current input states, while the forgetting gate determines how many states are discarded. The output gate regulates the number of internal states revealed to the cell for the next time step and the upper layer. The LSTM prediction process in this study follows these steps:

1. Identify input and output components.
2. Normalize the data.

3. Split the data into training, validation, and testing sets. Training data is used to learn unknown patterns, validation data ensures the network's appropriateness, and testing data is for forecasting with the trained model.
4. Experiment with the number of nodes in the hidden layer and the delay time.
5. Train the model.
6. Validate the model.
7. Generate forecasts.
8. Denormalize the data.
9. Divide the data into training, validation, and testing sets. The training data is utilized to learn unknown patterns, while the validation data guarantees that the created network is appropriate. The testing data is utilized for forecasting using the trained and tested model.

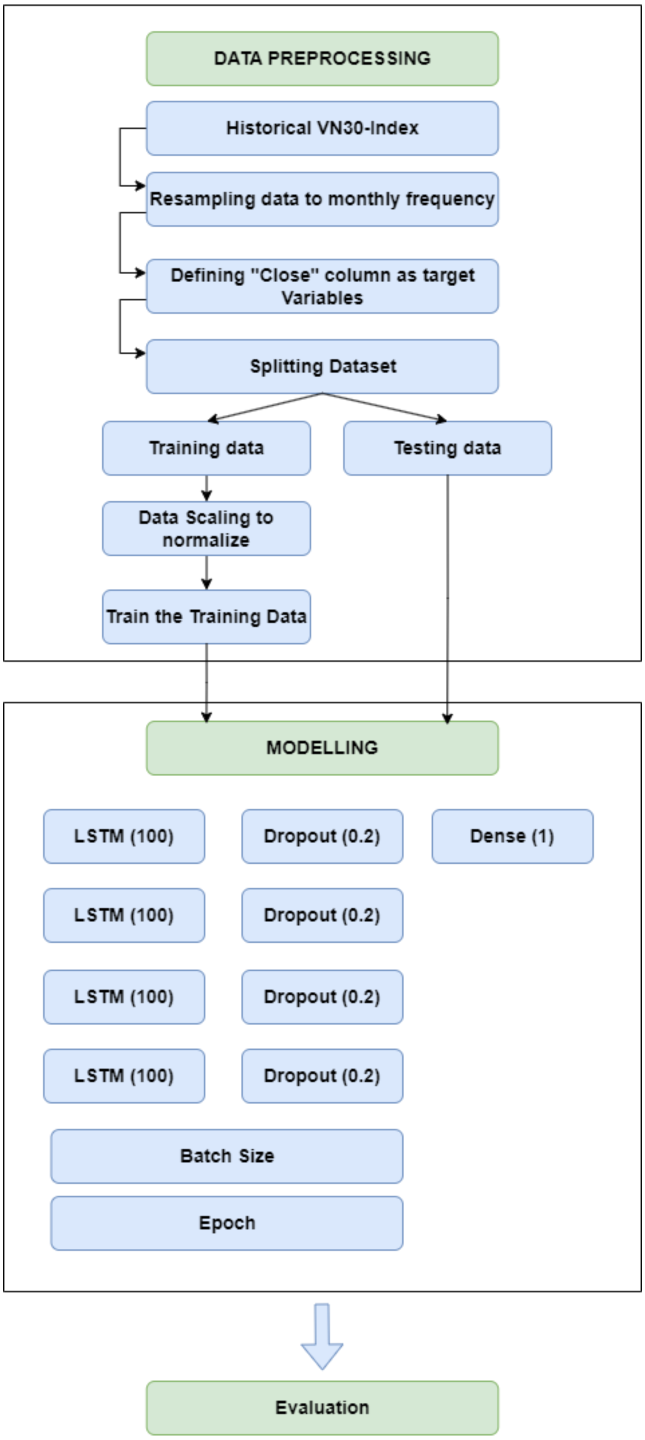

Figure 4. LSTM Framework structure

Furthermore, for the LSTM approach, we have designed a framework encompassing data preprocessing, model development, and evaluation, as demonstrated in Figure 4.

3. RESULTS AND DISCUSSION

3.1. SARIMA

The first step of the SARIMA model involves converting the data into a time series using a time series function. This allows us to check the stationarity of the data by examining a visualization of balance statistics. At this stage, we tested the data stationarity using the ADF test. The rolling statistics and graphs in this study reveal a p-value of < 0.05, which means we reject H0 (null hypothesis) of non-stationary data. The result of the ADF test is p-value = $0.587659 > 0.05$ showed that the null hypothesis (H0) could not be rejected.

To address the increasing levels and trends in the time series, we applied the box-cox transformation method to the time series data. This transformation allowed us to achieve the rolling average of the series by taking the input data over the past 12 months and providing the average consumption values at each point in the series. Subsequently, we partitioned the recorded time series data into two sets: 80% as training data and 20% as test data. To identify and separate seasonality and trends, we utilized a segmentation process.

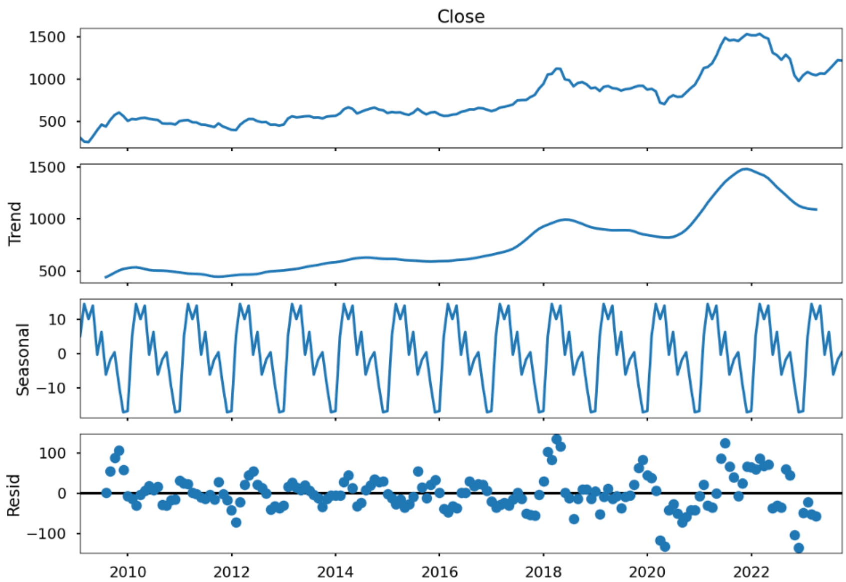

Figure 5. Decompose close index of VN30

The next step in our analysis involves distinguishing time series data that has been transformed using the box-cox transformation. This process allows us to determine the I (Integrated) value for the SARIMA model and evaluate whether the data series is stationary or not. Specifically, we have checked the p-values and looked for values below 0.05.

We observed that first-order differencing is necessary. The results of our ADF test, with a 12-month lag, yielded a p-value of 0.000399, which is less than the significance level of 0.05. This indicates that the series has achieved stability at the first difference and exhibits annual seasonality. From these findings, we discerned that the use of a ARIMA model would be inappropriate for analyzing this seasonally adjusted data. Consequently, we opted for the

SARIMA (Seasonal ARIMA) model. By meticulously evaluating the stability of the data, we ensure that our model is robust and capable of providing accurate forecasts

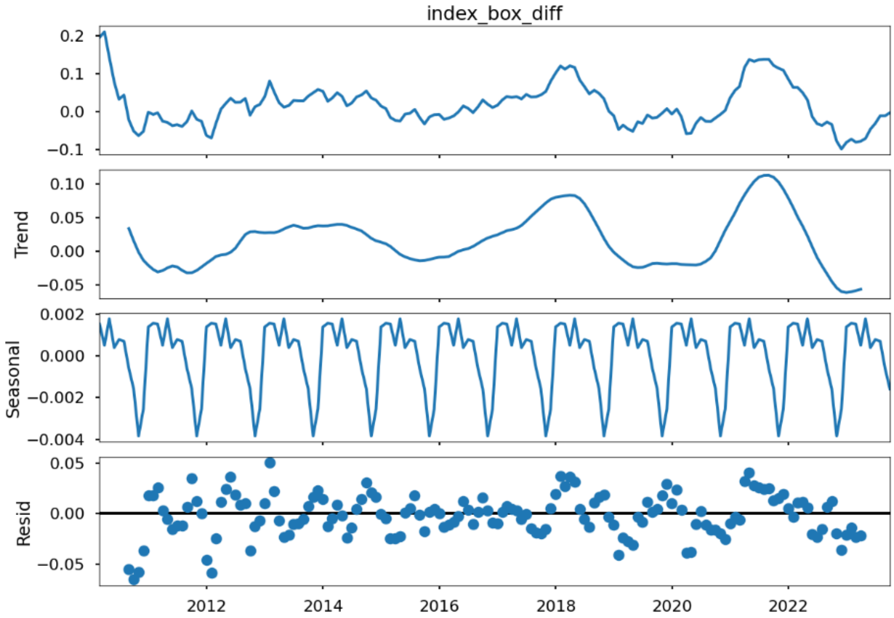

Figure 6. Decompose differential close index of VN30

Furthermore, the PACF value is determined, namely autoregressive (AR), and the ACF value, which is the moving average (MA), is the value of p, q.

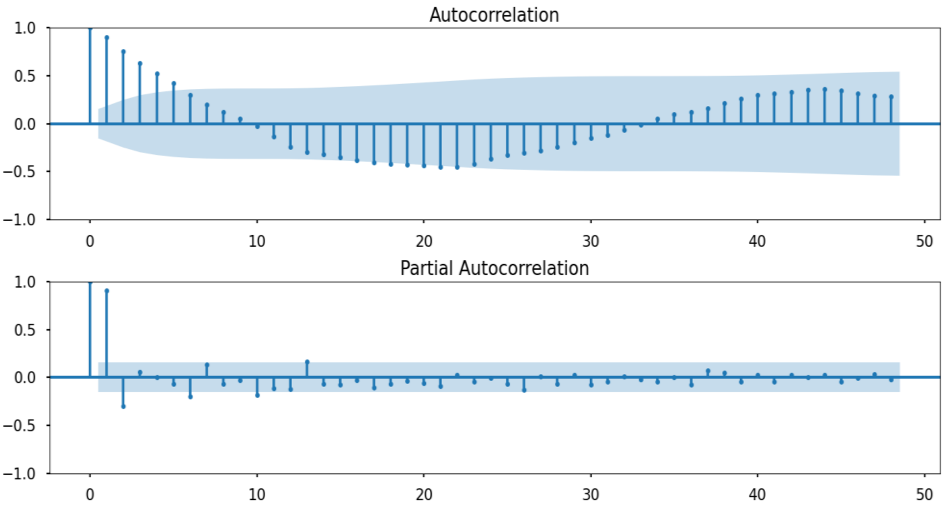

Figure 7. ACF and PACF to estimate p, q and P, Q

The determination of SARIMA models can be estimated with p, q, P, and Q, where P and Q respectively represent the order of the seasonal AR and MA components. Based on the autocorrelation plot (Figure 7), we see that ACF = 0 immediately after lag 5; therefore, the maximum value of q is 5. Similarly, from the partial autocorrelation plot (Figure 7), we see PAC = 0 immediately after lag 2; therefore, the maximum value of p is 2. Similarly, we get the maximum value of Q is 5 and the maximum value of P is 2.

To determine the best (p, d, q)(P, D, Q) values, we select the smallest Akaike Information Criterion (AIC),Bayesian information criterion (BIC) and Hannan–Quinn information criterion (HQIC) value, which corresponds to the number of differencing (one-time differencing). Once the best (p,d,q) and (P,D,Q) values are identified, we proceed to model by inputting the order value.

Table 1. Results of 4 SARIMA forecasting models

Model	**AIC**	**BIC**	**HQIC**
SARIMA(2, 1, 2)(0, 1, 1, 12)	-683.004491	-665.845617	-676.032493
SARIMA(2, 1, 2)(1, 1, 1, 12)	-682.996723	-662.978036	-674.862725
SARIMA(2, 1, 3)(0, 1, 1, 12)	-682.544767	-662.526080	-674.410770
SARIMA(2, 1, 2)(0, 1, 3, 12)	-681.899306	-659.020806	-672.603308
SARIMA(0, 1, 1)(0, 1, 1, 12)	-681.771255	-673.191818	-678.285256

In cases where all AIC, BIC, HQIC indicators are negative, we will choose the one with the highest value. In this article, we will select the SARIMA(2, 1, 2)(0, 1, 3, 12) model, although this model does not have the highest AIC, it has the highest BIC and HQIC.

Below are the results of forecasting VN30 data by month using the SARIMA(2, 1, 2)(0, 1, 3, 12) model.

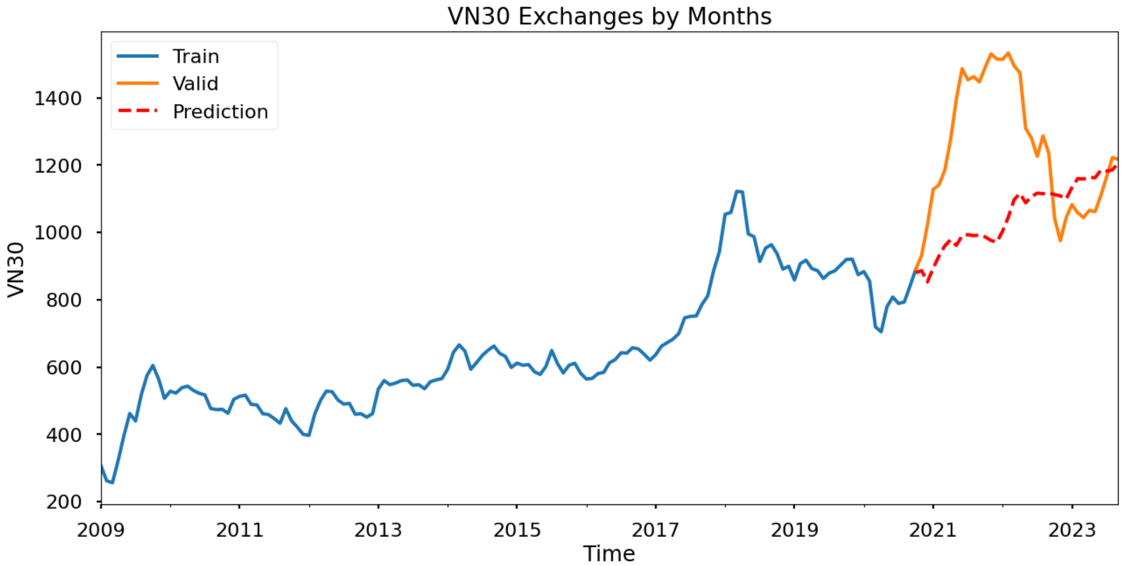

Figure 8. Forecasting result for model SARIMA(2, 1, 2)(0, 1, 3, 12)

From the results, it is evident that the SARIMA(2, 1, 2)(0, 1, 3, 12) model generally accurately forecasts the upward trend of the market. However, it falls short in precisely capturing specific fluctuations, such as the remarkable growth experienced in 2021 or the substantial market correction that occurred in 2022.

Although this model does not perform exceptionally well in forecasting tasks, it is still capable of capturing volatility and indicating trends and seasonality. And the SARIMA(2, 1, 2)(0, 1, 3, 12) result accuracy metric values displayed in MAPE and RMSE units (Table 2).

Table 2. MAPE and RMSE of model SARIMA(2, 1, 2)(0, 1, 3, 12)

MAPE	RMSE
17.52%	298.2714

3.2. LSTM

To forecast the VN30 Index using the LSTM methodology, we allocated 80% of our dataset for training purposes, with the remaining 20% set aside for validation and testing. The initial procedure involves data normalization using the MinMaxScaler feature. This step creates an equal number of training data objects as present in the training set, scales the data accordingly, and converts it into a three-dimensional format.

Subsequently, the model construction phase begins, incorporating four LSTM layers, each containing 100 neurons. This is complemented by four Dropout layers with a dropout rate of 0.2, and a concluding Dense layer. The training approach adopts an iterative function, analyzing each batch of the first ten data points (from index 0 to 9) and utilizing the subsequent data point (index 10) as the target for prediction.

The model is then compiled using the Adam optimization algorithm. Training is conducted over 50 epochs with a batch size of 25. Following the training phase, the model undergoes a similar process with the test data, involving the creation of test data objects corresponding to the number of test entries, scaling, and reshaping the data into a three-dimensional structure.

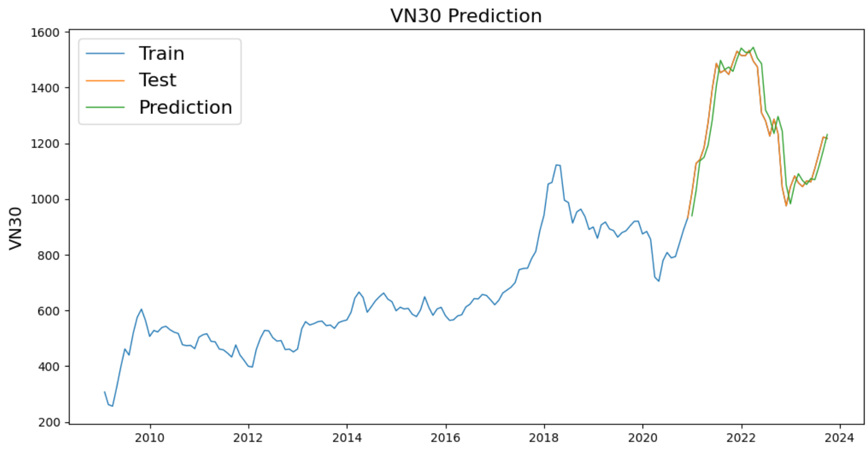

Figure 9. Prediction result for LSTM model

Figure 9 demonstrates the effectiveness of the Long Short-Term Memory (LSTM) method in accurately capturing and following the variations in the VN30-Index, exhibiting a notably enhanced performance compared to the previously discussed SARIMA model. Additionally, Table 3 highlights the effectiveness of the LSTM model, as evidenced by its Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) metrics.

Table 3. MAPE and RMSE of LSTM model

MAPE	RMSE
4.25%	67.485

4. CONCLUSION

This study aims to compare the accuracy of the SARIMA and LSTM methods in forecasting time series data models. The results indicate that the LSTM method outperforms the SARIMA method in both accuracy and visualization. While SARIMA provides a general trend of the data and its seasonality, LSTM can more effectively read data movements and track the fluctuations of the VN30-Index. Additionally, LSTM allows for adjustments in the number of neurons, batch size, and epoch values in both the training and testing phases to enhance accuracy. Future research could further improve accuracy by increasing these parameters in each layer, considering the high volatility in cryptocurrencies. Overall, this study contributes to understanding the efficacy of different methods in time series forecasting and offers insights for future research in this field.

However, for the Vietnamese stock market, the forecasted indices on certain trading days can significantly differ from reality, largely influenced by factors such as investor psychology, impacts from other stock markets, and policy changes. Investors should combine forecast results with technical analysis and regular observation for a more accurate understanding of market fluctuations. Additionally, to make informed investment decisions for individual stocks, investors might consider using the Capital Asset Pricing Model (CAPM).

SARIMA and LSTM models serve to simulate past market behavior, providing a basis for future forecasts. In Vietnam, the significance of forecasting models is not yet fully realized, as policymakers tend to use administrative measures to adjust trading margins with the aim of stabilizing the stock market, rather than addressing the main causes of market downturns. Therefore, the government should implement and consistently apply a system of policies to avoid strong market shocks and restore investor confidence.

The results of the SARIMA and LSTM models are primarily for reference purposes. However, it can be affirmed that within the scope of this study, both SARIMA and LSTM models are effective for forecasting market trends.

REFERENCES

[1]. Phan, T. T. D., & Luong, N. T. D. (2020), "Áp dụng mô hình ARIMA trong Dự báo biến động chỉ số VN30 và lợi suất chỉ số VN30," *Journal of Economics and Development*, 276(6), ISSN 1859 - 0012.

[2]. Seabe, P. L., Moutsinga, C. R. B., & Pindza, E. (2023), "Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach," *Fractal Fract*, 7(2), 203, doi: https://doi.org/10.3390/fractalfract7020203.

[3]. Zahara, S., Sugianto, & Ilmiddaviq, M. B. (2020), "Consumer price index prediction using Long Short Term Memory (LSTM) based cloud computing," *Journal of Physics: Conference Series*, 1456(1), 012022, IOP Publishing, doi: https://doi.org/10.1088/1742-6596/1456/1/012022

[4]. S. Sa'adah and M. S. Wibowo (2020), "Prediction of Gross Domestic Product (GDP) in Indonesia Using Deep Learning Algorithm," *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, pp. 32-36, doi: 10.1109/ISRITI51436.2020.9315519.

[5]. J. Sun (2019), "Forecasting COVID-19 pandemic in Alberta, Canada using modified ARIMA models," *Comput. Methods Programs Biomed. Updat.*, vol. 1, no. April, p. 100029, 2021, doi: 10.1016/j.cmpbup.2021.100029.

[6]. Pasak, S., & Jayadi, R. (2023), "Investment Decision on Cryptocurrency: Comparing Prediction Performance Using ARIMA and LSTM," *Journal of Information Systems and Informatics*, 5(2), 407, doi: https://doi.org/10.51519/journalisi.v5i2.473

[7]. Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018), "Forecasting of demand using ARIMA model," *International Journal of Engineering Business Management*, doi: https://doi.org/10.1177/1847979018808673

[8]. Kobiela, D., Krefta, D., Król, W., & Weichbroth, P. (2022), "ARIMA vs LSTM on NASDAQ stock exchange data," *Procedia Computer Science*, 207, 3830-3839< doi: https://doi.org/10.1016/j.procs.2022.09.445

[9]. Ngọc, L. N. B., Thông, L. Q., & Hòa, T. A. (2018), "Mô hình dự báo giá tôm sú xuất khẩu Việt Nam," *CTU Journal of Innovation and Sustainable Development*, 54(6D), 188-195. doi: https://doi.org/10.22144/ctu.jvn.2018.111

[10]. Martinez, E. Z., da Silva, E. A. S., & Dal Fabbro, A. L. (2011), "A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo", Brazil,

Revista da Sociedade Brasileira de Medicina Tropical, 44(4), 436-440, doi: https://doi.org/10.1590/s0037-86822011000400007

[11]. Yadav, A., Jha, C. K., & Sharan, A. (2020), “Optimizing LSTM for time series prediction in the Indian stock market,” *Procedia Computer Science*, 167, 2091-2100, doi: https://doi.org/10.1016/j.procs.2020.03.257

The code of this paper can be found here.