Accounting and Auditing with Application

www.aaa.reapress.com

Acc. Aud. Appl. Vol. 2, No. 2 (2025) 109-121.

Paper Type: Original Article

Intelligent Stock Price Prediction Using LSTM, GRU, ARIMA, and ARIMAX Models: Analysis and Performance Comparison

Seyyede Bita Amiri[1], **Arefeh Amidian**[1,*], **Zohre Fasihfar**[1]

[1] Department of Computer Science, Hakim Sabzevari University, Sabzevar, Iran; bitaamiri.1381@gmail.com; arefehamidiyan@yahoo.com; zfasihfar@gmail.com.

Citation:

Received: 19 June 2024 *Revised: 14 September 2024* *Accepted: 27 November 2024*	Amiri, S. B., Amidian, A., & Fasihfar, Z. (2025). Intelligent stock price prediction using LSTM, GRU, ARIMA, and ARIMAX models: Analysis and performance comparison. *Accounting and auditing with application, 2*(2), 109-121.

Abstract

This study examines and compares the performance of four-time series forecasting models, including ARIMA, ARIMAX, LSTM, and GRU, in forecasting the stock price of Iran Export Bank over 16 years (2009-2025). The data were prepared for modeling after performing preprocessing steps such as normalization. In the modeling section, the classical ARIMA models and the improved version of ARIMAX, which also consider exogenous variables (such as trading volume, moving average, and volatility), were used along with deep learning-based Recurrent Neural Networks (RNNs), namely LSTM and GRU. The results showed that the deep learning models LSTM and GRU performed significantly better than the classical models. Among the classical models, ARIMAX performed significantly better in forecasting than ARIMA, which had very poor performance. The LSTM model provided the most accurate forecasts and was able to model more than 98.67 percent of the data changes. The GRU model also performed close to LSTM, approximately 98.61, but its accuracy was slightly lower than LSTM. The results show that deep learning models, especially LSTM, perform better than classical models in simulating nonlinear patterns and long-term dependencies in financial data. Overall, deep learning-based models, especially LSTM, are powerful tools for predicting complex time series and can play an important role in investment decisions and analyzing stock market trends.

Keywords: Time series, Deep learning, LSTM, GRU, ARIMAX.

1 | Introduction

In recent decades, the complexity and volatility of financial markets have attracted the attention of researchers and economic practitioners to develop more accurate stock price forecasting methods [1]. The ability to

Corresponding Author: arefehamidiyan@yahoo.com

https://doi.org/10.22105/aaa.v2i2.67

forecast stock prices allows investors to adopt intelligent strategies for making buying and selling decisions and minimize their investment risk [2]. In this regard, advanced data analysis methods, including deep learning and time series models, have become essential tools in financial market analysis. Recurrent Neural Networks (RNNs) such as GRU and LSTM are widely used in time series forecasting due to their ability to identify long-term and complex patterns in financial data [3]. These models help identify future trends by learning from the past behavior of data.

Along with these methods, the classic ARIMA model is still considered one of the popular tools for time series analysis due to its linear structure [4], [5]. This study compares the performance of advanced LSTM and GRU models with ARIMA and ARIMAX models and examines the effectiveness of each in stock price forecasting. This research aims to identify the best-performing algorithm for more accurate prediction of stock prices in the future and provide a framework for optimal financial decision-making**.**

2 | Literature Review

Zhang et al. examined the performance of ARIMA and LSTM models in predicting the prevalence of Hand, Foot, and Mouth Disease (HFMD) in Ningbo, China, between 2014 and 2017 and investigated the improvement of prediction accuracy in environmental factors by adding weather variables. The results showed that LSTM models, especially its multivariate version, performed better than ARIMA and could better recognize nonlinear patterns [6].

Ulyah et al. examined the impact of the 2019 Indonesian presidential election on the composite index IDX (JKSE) and SRTG stock returns using ARIMAX and ARIMA-GARCH models. The results showed that the exogenous variable related to the election did not statistically affect the returns of this index and stocks. Therefore, the ARIMA (3,0,2) model performed better for JKSE because it better modeled the time series changes without external variables.

In some cases, simpler models such as ARIMA can provide better performance [7]. Khaled et al. [8] worked on stock market forecasting using LSTM and GRU neural networks. They compared these architectures in stacked and bidirectional forms for short-term and long-term forecasting. They showed that Stacked LSTM (SLSTM) performs better and more accurately than GRU and other models [8].

Jiang [9] investigated stock market price forecasting using deep learning techniques, examined data such as historical prices, news, and economic indicators and evaluated various models, including RNNs. They finally concluded that deep learning models outperform classical methods such as ARIMA and machine learning algorithms such as SVM [9]. Hu et al. studied stock and currency market price prediction using deep learning techniques. They analyzed and evaluated models including CNN, LSTM, RNN, DNN, and reinforcement methods and found that combining LSTM models with DNN provides better predictions [10].

Sivasamy conducted a study in 2024 on analyzing and predicting gold prices using machine learning methods, including GRU, LSTM, HMM, and SARIMA. He divided 11,151 historical gold price data days into training and test sets (including the last 453 days). He evaluated the performance of the models using metrics such as RMSE and MAE. His results showed that the GRU model outperformed LSTM and ARIMA [11].

However, the findings of Biswas et al. in examining 12-month data of the Dhaka Stock Exchange (DSE) show that LSTM has higher prediction accuracy than other models, namely XGBoost, linear regression, and moving average. This result is consistent with the present study because, in this study, LSTM was evaluated as the most accurate model [12]. Siami-Namini et al. showed that LSTM is significantly more accurate, and the average error rate reduction is between 84% and 87%. Also, increasing the number of epochs did not significantly improve model performance [13].

Atsalakis and Valavanis investigated the prediction of stock market prices and indices using soft computing methods. Their findings show that determining the optimal model structure, including the number of layers and neurons, remains challenging and requires trial and error. Their research results also indicate that soft computing techniques (artificial neural networks, fuzzy systems, genetic algorithms, support vector machines, neuro-fuzzy systems) are more accurate in forecasting prices than traditional models such as linear regression and ARIMA [14].

3 | Data Preprocessing

The 16-year (2009-2025) dataset of Iran Export Bank (BSE) stocks was collected in CSV format, which displays the final and closing prices of the stocks in the 16-year time frame and was extracted from the Tehran Stock Exchange Technology Management website. This dataset contains 3795 records that store daily information about the stock price in different columns. The columns include dates and prices such as first, highest, lowest, closed, and trading amounts. The data preprocessing steps are as follows:

3.1 | Data Cleaning

The initial data review showed no duplicates or missing values in the dataset. Then, outliers were identified and managed, as these values can negatively impact the performance of machine learning models and statistical analyses, and a comprehensive review of outliers was conducted. Different thresholds were used in the Z-Score method [6], [15] to identify these data. In the first step, the threshold value was set to 3, which resulted in the identification of 86 outliers. Then, this value was increased to 4, which reduced the number of outliers identified to 39. Subsequently, the threshold value was set to 5, which resulted in only 2 outliers.

After identifying outliers, various approaches were examined to replace or remove them. One of the methods used was to replace outliers with the mean value of the dataset. Direct removal of these data was also tested. However, the results of these processes showed that any change in the dataset leads to a decrease in the accuracy of deep learning models. In contrast, when the data was not changed, that is, without removing or replacing outliers, the performance of the algorithms improved.

Accordingly, in this particular problem, preserving outliers and not making changes to the original dataset improves the accuracy of the models. This finding indicates that removing or replacing outliers sometimes does not improve the results.

3.2 | Data Scaling

Scaling operations were performed on the numerical features of the data to improve the performance of machine learning models and reduce the impact of different feature scales. This operation ensures that all features are in the same range, i.e. [0, 1], and the models are trained more accurately. For this purpose, the MinMaxScaler tool available in the Scikit-Learn library was used [2], [16], [17].

$$x_{\text{scaled}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}. \tag{1}$$

Numeric attributes, including close, open, high, low, and vol, were selected for scaling, and other columns, such as date, were excluded from the process due to their non-numeric nature or inappropriate scaling.

3.3 | Data Preparation with Augmented Dickey-Fuller Test for Classical Models

One of the main prerequisites for using ARIMA and ARIMAX models is that the data are stationary. The Augmented Dickey-Fuller (ADF) test is used to check the stationarity of the data. This test checks the presence or absence of a unit root in the time series. If the time series has a unit root, it is not stationary and requires transformation or differentiation for modeling.

If the p-value of the ADF test is less than a specified level (usually 0.05), the null hypothesis is rejected, and the data are considered stationary; otherwise, appropriate differentiation or transformations are required to achieve stationarity. This test showed that the time series becomes stationary after one differencing. That is, the data should be differencing once before entering ARIMA and ARIMAX; however, due to the improvement of the performance of non-differentiation, differencing was not performed in this study [6], [7], [18].

3.4 | Data Preparation for Deep Learning Models

After the data cleaning and scaling steps were performed, the time series data were prepared for training the deep learning models. The data from 2009 to 2021 were considered the training set, and the data from 2022 to 2025 were considered the test set**.**

The data was converted into sequential time sequences to use recurrent networks. Since different choices of the number of input time steps led to different results in the performance of the LSTM and GRU models, the results obtained were used as the basis for the optimal selection of this parameter for tuning the models. Accordingly, by examining different values of the time step starting from 22 and increasing to 35, 90, and 170, respectively, according to *Table 1*, the performance of the models was evaluated. Also, to be compatible with the input structure of these models, the data were rearranged into a three-dimensional format [samples,time_steps, features].

4 | Methods

4.1 | AutoRegressive Integrated Moving Average (ARIMA) Model Method

The ARIMA model is a classic method for forecasting and analyzing time series, which consists of three main components, known as ARIMA for short:

(Auto Regressive) AR: The autoregressive part of the model states that the future value of the time series depends on its past values. In the AR model, each data point depends on previous data points, representing this parameter by p.

(Integrated) I: The part related to transforming the data into a stationary state. For the ARIMA model to be applied to the data, the data must be stationary. The value d indicates the number of differentiations required to make the data stationary.

(Moving Average) MA: The moving average part of the model that considers the noise or errors in the past forecasts that affect the current value, and its parameter is q.

By combining these three components, the ARIMA model forecasts the time series by considering past correlations and the overall trend of the data [19].

4.2 | AutoRegressive Integrated Moving Average with Exogenous Variables (ARIMAX) Model Method

The ARIMAX model is an improved version of ARIMA that includes exogenous variables and the dependent variable (closing stock price). This allows the impact of key market factors on the time series to be considered, while ARIMA predicts the same variable using only past values. In this study, the following auxiliary variables were considered as inputs to the ARIMAX model:

Trading volume: the amount of shares traded per day.

10-day moving average: an indicator that shows the medium-term price trend and helps reduce price noise [10]. Volatility: a key factor in measuring risk that reflects sudden price changes [20].

4.3 | Model Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU)

Conventional RNNs have difficulty learning long-term dependencies because the gradient magnitude decreases drastically with increasing time steps, hindering effective learning. To overcome this challenge, more advanced architectures such as LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) were introduced; the differences between their architectures are illustrated in *Fig. 1* [19].

LSTM consists of three main gates: the Forget gate, which determines which information is deleted from memory; the Input gate, which determines what new information is added to memory; and the Output gate, which controls what data is published as output. Due to its complex structure, this architecture can learn long-term dependencies in sequential data, such as time series forecasting.

On the other hand, GRU has a simpler structure. It consists of only two gates: the Update gate, which combines new information with memory, and the Reset gate, which determines how much previous information is forgotten. This architecture is faster due to its lower complexity and, in some cases, performs similarly or even better than LSTM.

Considering the above features, the choice between these two architectures in *Fig. 1* depends on the problem's needs. LSTM is more suitable for data with long and complex time dependencies because its three-gate structure helps to retain important information over longer periods. Due to its simpler structure, GRU performs better in problems with less data or requiring faster processing and usually provides similar results to LSTM.

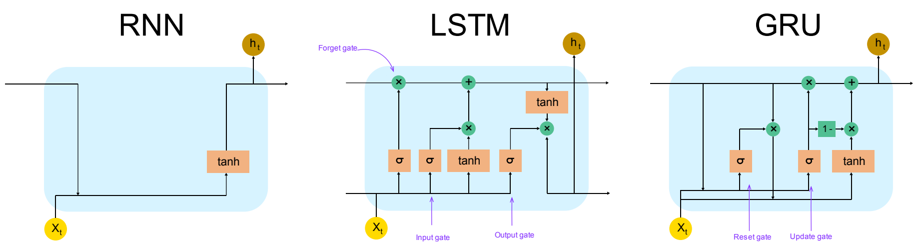

Fig. 1. GRU, LSTM, and RNN architecture.

5 | Training and Evaluation

The Google Colab platform and various libraries such as Pandas, NumPy, Matplotlib, Statsmodels, Scikit-Learn, and Keras were used to run the forecasting models. The time series data were divided into two training and test sections to train and evaluate the models. Data from 2009 to early 2022 (13 years) were used to train the models, and data from 2022 to early 2025 (3 years) were used to evaluate the model performance.

In this division, about 81% of the data was allocated to training, and 19% was considered for testing. After training the model, predictions were made for the test data, and the model performance was evaluated using the following criteria:

Mean Squared Error (MSE): The mean squared difference between the actual and predicted values [16].

Mean Absolute Error (MAE): The mean absolute difference between the actual and predicted values [6].

Coefficient of determination (R^2): The extent to which the model explains the variance in the data [21].

5.1 | Training and Evaluating the ARIMA Model

5.1.1 | Setting up and training the ARIMA model

As specified in the preprocessing section, the parameter d was set to zero. The autoregression lags, i.e., the parameter p in this model, was set to 3, and the parameter of the moving average terms, i.e., q, was set to 2.

The ARIMA (3,0,2) model was set [7], although the model setting was done as ARIMA (5,1,0), but the result was not desirable.

5.1.2 | Forecasting and evaluating the ARIMA model

The results of the ARIMA model evaluation in *Fig. 2* (yellow line) indicate this model's stable and uniform behavior, especially in intervals where real data fluctuates greatly. This indicates the model's weakness in adapting to time series changes.

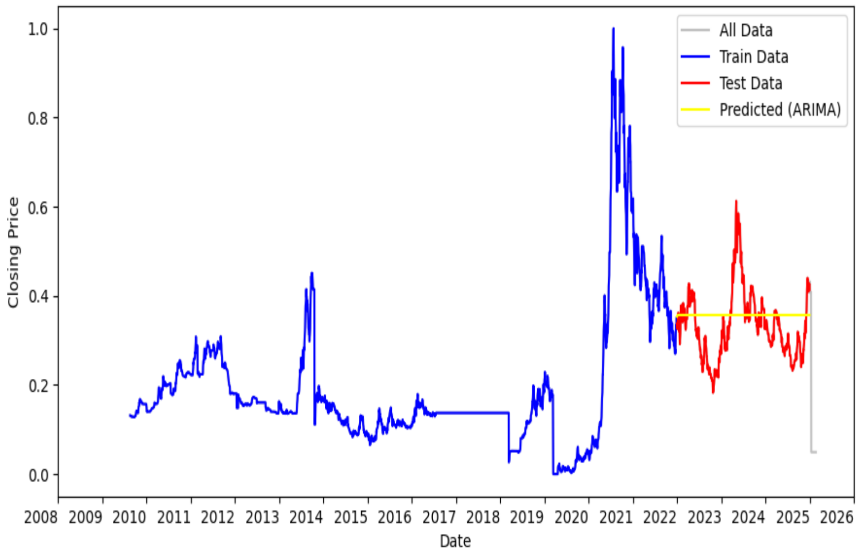

Fig. 2. ARIMA forecast chart.

The results of the criteria for evaluating the accuracy of the ARIMA model are as follows:

I. MSE: 0.006

II. MAE: 0.064

III. R^2: -0.109

A low MSE value indicates that the model has predictions close to the actual values at some points, but this does not mean that the model performs well. Like MSE, the MAE value also indicates that the predictions do not deviate from the actual data.

MAE is usually more tolerable than MSE, but this value is still considered high. However, even if these two measures have small values, a negative R^2 value indicates the model's inability to explain and predict the behavior of the data. Typically, R^2 is between 0 and 1, and negative values mean the model has performed even worse than a simple prediction (such as the mean).

5.2 | Training and Evaluation of the ARIMAX (Auto Regressive Integrated Moving Average with Exogenous Variables) Model

5.2.1 | Setting and training the ARIMAX model

The ARIMAX model is tuned and evaluated in this section to predict stock closing prices. First, new features are created to help the model better predict the data. These features include the 10-day moving average of the closing price [10], which indicates the price trend changes, the standard deviation of the closing price in a 10-day window [12], [22] which indicates the amount of volatility, and the trading volume, which is added as a feature.

After the feature engineering process and data partitioning, the ARIMAX (3,0,2) model is fitted to the training set using external variables [7].

5.2.2 | Forecasting and evaluating the ARIMAX model

These results show that the ARIMAX model could provide forecasts with negligible errors and explain a significant portion of the data variance. *Fig. 3* shows the forecast graph of the ARIMAX model (yellow line). Compared with the ARIMA model, the behavior of the ARIMAX model is better consistent with changes in the actual data, indicating an improvement in the model's ability to identify time series patterns.

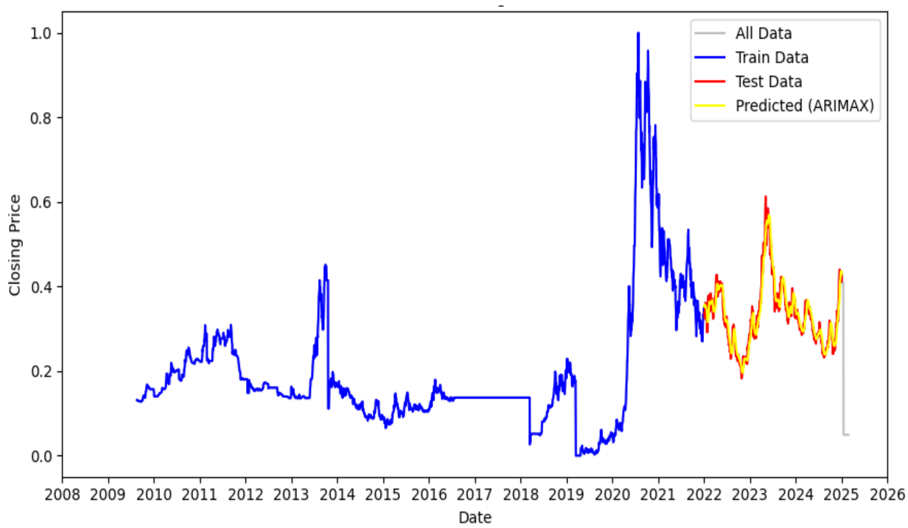

Fig. 3. ARIMAX forecast chart.

After fitting the ARIMAX model on the training set, the model predictions were made for the test set, and the performance results of the ARIMAX model show a significant improvement in prediction compared to the ARIMA model:

I. MSE: 0.000

II. MAE: 0.016

III. R^2: 0.931

5.3 | Training and Evaluating the LSTM Model

5.3.1 | Setting up and training the LSTM model

The designed LSTM model has two recurrent layers with 100 neural units that use the Dropout mechanism to reduce overfitting, the value of which is set to 0.1 in both recurrent layers. A Dense layer with 64 neurons and a ReLU activation function is added to process more features.

The model was optimized using the mean square error (MSE) function, and the Adam optimization algorithm, which had a learning rate of 0.001, was selected to improve the learning process. The model was trained using training data over 200 epochs and with a batch size of 32 samples. To prevent overfitting, the Early Stopping mechanism was set to patience = 20, so the training process would stop early if performance did not improve.

5.3.2 | LSTM model prediction and evaluation

The test data was predicted using the model after training the LSTM model with the specified settings. The graph in *Fig. 4* compares the actual values and the values predicted by the LSTM model. As can be seen in the graph, the yellow line (model predictions) matches the red line (actual values) very closely, demonstrating the model's ability to detect nonlinear changes and complex time dependencies.

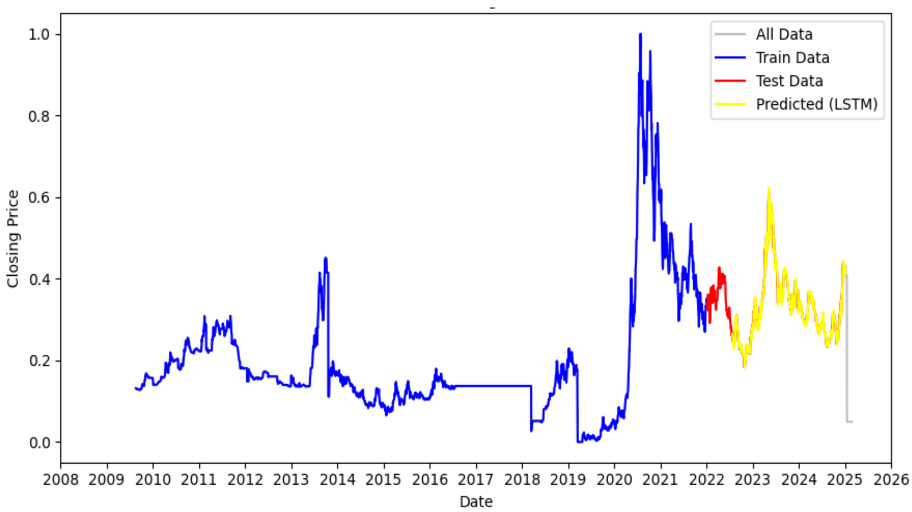

Fig. 4. LSTM forecast chart.

The results of the criteria for evaluating the accuracy of the LSTM model are as follows:

I. MSE: 0.00009

II. MAE: 0.00672

III. R^2: 0.98674

A very small MSE value indicates a high accuracy of the model. A low MAE value also indicates a high accuracy of the model in predicting the exact value of prices. A high R^2 value close to 1 indicates that the model was able to explain about 98.67% of the variation in the actual data, which indicates a strong performance of the model in identifying trends and patterns in time series data.

5.4 | Training and Evaluating the GRU Model

5.4.1 | Setting up and training the GRU model

The improved GRU model is designed with three recurrent layers: the first layer contains 128 neural units to extract complex temporal features, the second layer has 64 neural units with sequential output to retain more details from the input data, and the third layer contains 32 neural units to reduce dimensions and extract final features that learn the complexity of the patterns, respectively.

In all three GRU layers, the dropout mechanism with rates of 0.2 and 0.1 was used to reduce overfitting. Finally, a dense layer with one neuron was added to the model to predict the stock's closing price for the next day.

The model was optimized using the Mean Square Error (MSE) function, and the Adam optimization algorithm, which had a learning rate of 0.0005, was used to improve the learning process and reduce the prediction error. The model was trained over 200 epochs with a batch size of 32 samples. The Early Stopping mechanism was set to 20 for the patience parameter to prevent overfitting.

5.4.2 | Prediction and evaluation of the GRU model

After training the GRU model, the test data was predicted using the model. The graph in *Fig. 5* compares the actual values and the values predicted by the GRU model, where the yellow line (model predictions) matches well with the red line (actual values). The model follows the overall trend of price changes with very good accuracy. Minor deviations are observed at some swing points, especially at sharp price peaks and troughs. However, these insignificant differences indicate the model's performance in modeling price changes.

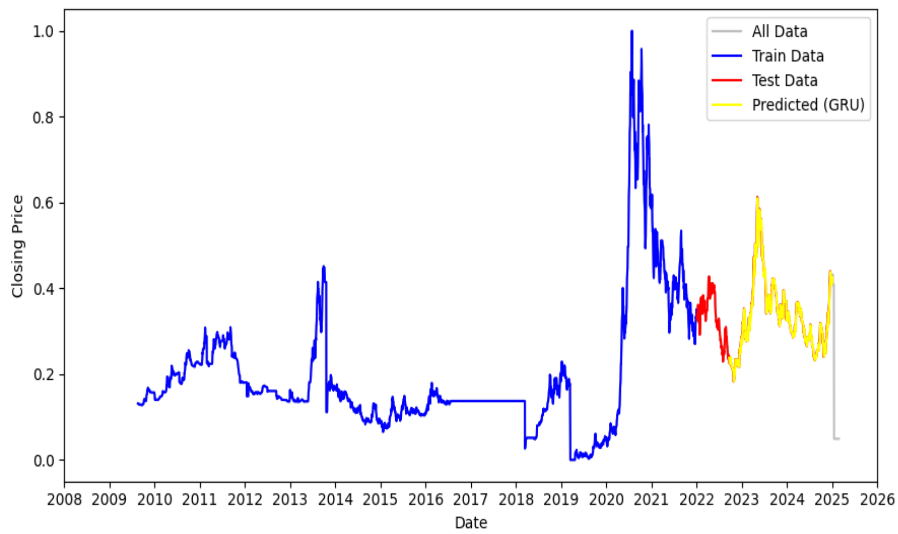

Fig. 5. GRU forecast chart.

The results of the criteria for evaluating the accuracy of the GRU model are as follows:

I. MSE: 0.0000927

II. MAE: 0.0070912

III. R^2: 0.98618

The low MSE and high R^2 values indicate that the model accurately predicts the test data. Although there are small differences between the actual and predicted values in some extreme price fluctuations, these differences are limited. These results indicate the high capability of the GRU model, which was able to explain more than 98.61% of the stock price changes.

6 | Comparison and Analysis of Results

This section investigated and analyzed the performance of four models, ARIMA, ARIMAX, GRU, and LSTM, based on MSE, MAE, and R^2 criteria. The results show that deep learning models, especially GRU and LSTM, perform much better than traditional models, such as ARIMA and ARIMAX, predicting complex data with nonlinear patterns.

6.1 | Performance Analysis of ARIMA and ARIMAX Models

The ARIMA model, one of the classic time series forecasting models, failed to model the fluctuations in the time series in this study and for these data. Although this model is well capable of modeling simple and linear trends, the results obtained from its evaluation indicate its poor performance in predicting more complex data. The evaluation values show that the ARIMA model cannot predict accurate values and simulate the overall behavior of the data, and it has performed very poorly. Its forecasting graph (*Fig. 2*) also shows the model's stable and uniform behavior in conditions where the real data has many fluctuations, clearly showing the model's weakness even in the smallest forecast.

The ARIMAX model provided more accurate forecasts than the ARIMA model, which did nothing. In particular, by adding variables such as trading volume, moving average, and volatility as exogenous features, this model could simulate more complex changes and nonlinear patterns in the data well. The evaluation results of the ARIMAX model show the high accuracy of this model in forecasting. It has been able to explain a large part of the variance in the data and provide favorable forecasts.

The ARIMAX forecasting graph (*Fig. 3*) also shows a very good agreement of the forecasts with the actual data, which clearly shows that the model has been able to simulate the patterns in the data well. In general, the ARIMAX model can simulate data changes more accurately than the ARIMA model, which depends on the external variables added to the model.

Given that the Dickey-Fuller test showed that the data were not stationary, the data should be differentiated before entering ARIMA and ARIMAX. However, experimental results showed that this model with a value of d equal to zero provides the best results compared to when differentiation is performed.

6.2 | Performance analysis of GRU and LSTM models based on the coefficient of determination (R^2) at different time steps

The results presented in *Table 1* indicate the performance of the GRU and LSTM models in predicting time series based on the coefficient of determination (R^2) at different time steps. Both models provide acceptable performance, but the LSTM model performed better than the GRU at all time steps examined.

With increasing time steps, the improvement trend of the coefficient of determination is evident for both models, but there are fluctuations at some points so that the R^2 value for GRU has increased from 0.98320 at time step 22, which is one working month [1] to 0.98618 at time step 170. This trend is also observed for LSTM, so the R^2 value has increased from 0.98409 at time step 22 [1] to 0.98674 at time step 130. Experimental results show that the performance of the LSTM model has improved at higher time steps compared to the reported values.

These results indicate that the LSTM model performs better in learning long-term dependencies, which is expected given its specific architecture in memory management and preventing information forgetting. Overall, this study's findings confirm that in time series forecasting applications where long-term dependencies need to be analyzed, the LSTM model can be a more suitable choice than the GRU. However, the LSTM's higher computational complexity may lead to increased training time and computational costs, which require further investigation in different situations.

Table 1. Performance comparison of GRU and LSTM with different time steps.

Time Step	R^2 (LSTM)	R^2 (GRU)
22	0.98409	0.98320
35	0.98449	0.98402
90	0.98505	0.98468
100	0.98619	0.98510
120	0.98640	0.98433
130	0.98674	0.98519
140	0.98557	0.98512
150	0.98640	0.98570
170	0.98628	0.98618

The LSTM model, one of the more advanced recurrent neural network models, showed the best performance among all the models and was able to simulate the long-term and nonlinear dependencies in the time series data well and provide accurate predictions. The results obtained from the evaluation of the LSTM model showed that the model has a very high accuracy in predicting prices and simulating data changes and was able to explain about 98.67% of the changes in real data. The LSTM prediction graph (*Fig. 4*) clearly shows a very close match of the predictions with the real data.

The GRU model, one of the advanced deep learning models, could simulate the complex features and long-term time dependencies in time series data well. Unlike ARIMA and ARIMAX, the GRU model is able to simulate more complex nonlinear patterns, and the results of the GRU model evaluation indicate that the model has a very high accuracy in data changes. These results show that the GRU model could provide accurate forecasts with very little deviation from the actual data and could simulate the overall trend of price changes well. In addition, the GRU model also showed good performance in simulating price fluctuations, with small differences between the forecasts and the actual data being observed at the peak and trough points of prices.

Increasing the number of epochs improves the model slightly, but from about 160 onwards, it has no effect on the accuracy of the LSTM and GRU models and leads to their random behavior. With the activation of early stopping, the training of the LSTM model stops at about 120 to 150 and in the GRU at about 80 to 100, and continuing the training after this range does not have a positive effect on the models' performance.

6.3 | Analysis and Comparison of the Performance of the 4 Models

The table below compares the performance of these four models based on MSE, MAE, and R^2 metrics. The ARIMAX model has better accuracy than ARIMA. Yet, the GRU and LSTM deep learning models still perform better in predicting complex data than the classical models, especially the LSTM, which provides the highest accuracy.

Table 2. Evaluation of the prediction error of the models.

Model	MAE	MSE	R^2
ARIMA	0.064	0.006	-0.109
ARIMAX	0.016	0.000	0.931
LSTM	0.00672	0.00009	0.98674
GRU	0.0070912	0.0000927	0.98618

The results show that the LSTM model has the best accuracy with the lowest error rate and the highest R^2 value. The ARIMAX model has improved compared to ARIMA by adding external variables, but it still lags behind the deep learning models. The ARIMA model also has the weakest performance and cannot simulate nonlinear patterns. Overall, deep learning models are recommended for predicting complex data.

7 | Conclusion

The results of the performance evaluation of the four models ARIMA, ARIMAX, GRU, and LSTM in predicting time series data indicate the significant superiority of deep learning models, especially GRU and LSTM, over traditional models such as ARIMA and ARIMAX. The GRU and LSTM models could simulate complex and nonlinear dependencies of the data well and provide accurate predictions. Among these models, the LSTM model showed the best performance and could predict more than 98.67 percent of the data changes with high accuracy.

The GRU model also provided acceptable results with the ability to simulate long-term time dependencies and nonlinear complexities. On the other hand, the ARIMA and ARIMAX models performed worse, especially when faced with complex data and high fluctuations. However, the ARIMAX model performed very well compared to ARIMA, which had an unacceptable performance when external variables were added. Finally, GRU and LSTM models are superior choices for predicting complex time series data.

Our suggestion for future research is to incorporate deep learning models and examine the impact of other economic variables (such as interest rates and inflation) as inputs to the models, which could help improve forecasting accuracy.

Acknowledgments

The authors sincerely express their gratitude to all individuals who contributed to this research. Special thanks to colleagues and financial analysts for their valuable insights and constructive feedback.

Author Contribution

Sayedeh Bita Amiri, Arefeh Amidian, and Zohre Fasihfar equally contributed to this research. Sayedeh Bita Amiri was responsible for conceptualization and methodology. Arefeh Amidian conducted data collection, preprocessing, and model implementation. Zohre Fasihfar contributed to result analysis, data interpretation, and manuscript writing. All authors reviewed and approved the final version of the manuscript.

Funding

This research did not receive any specific grant from public, commercial, or non-profit funding agencies.

Data Availability

The dataset used in this study is available upon reasonable request from the corresponding author. However, some restrictions may apply due to confidentiality agreements.

Conflicts of Interest

The authors declare that there are no conflicts of interest in this research.

References

[1] Shaygani, M., & Nourbakhsh, A. (2023). Predicting stock prices in the tehran stock exchange with artificial intelligence. *The second national conference on new developments in financial, economic and accounting studies*. (**In Persian**). Civilica. https://civilica.com/doc/2116487

[2] Abdi, N., MoradzadehFard, M., Ahmadzadeh, H., & Khoddam, M. (2024). A hybrid model for portfolio optimization based on stock price forecasting with LSTM recurrent neural network using multi-criteria decision making techniques and mean-CVaR. *International journal of finance & managerial accounting*, *9*(35), 1–14. http://www.ijfma.ir/article_22706.html

[3] Yaraqi, M., & Rabiei, A. (2021). Review and comparison of LSTM and GRU deep recurrent neural network algorithms in currency valuation time series data models. *Computer science*, *5*(4), 40-50. (**In Persian**). https://dor.isc.ac/dor/20.1001.1.2538161.1399.5.4.4.1

[4] Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, *33*(6), 497–505. https://doi.org/10.1016/j.omega.2004.07.024

[5] Gülmez, B. (2023). Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm. *Expert systems with applications*, *227*, 120346. https://doi.org/10.1016/j.eswa.2023.120346

[6] Zhang, R., Guo, Z., Meng, Y., Wang, S., Li, S., Niu, R., … & Li, Y. (2021). Comparison of ARIMA and LSTM in forecasting the incidence of HFMD combined and uncombined with exogenous meteorological variables in Ningbo, China. *International journal of environmental research and public health*, *18*(11), 1–14. https://doi.org/10.3390/ijerph18116174

[7] Ulyah, S. M. (2019). Forecasting index and stock returns by considering the effect of indonesia pre-presidential election 2019 using arimax and varx approches. *Journal of physics: Conference series* (Vol. 1277, p. 12053). IOP Publishing. https://iopscience.iop.org/article/10.1088/1742-6596/1277/1/012053/meta

[8] Althelaya, K. A., El-Alfy, E. S. M., & Mohammed, S. (2018). Stock market forecast using multivariate analysis with bidirectional and stacked (lstm, Gru). *2018 21st saudi computer society national computer conference (NCC)* (pp. 1–7). IEEE. https://doi.org/10.1109/NCG.2018.8593076

[9] Jiang, W. (2021). Applications of deep learning in stock market prediction: Recent progress. *Expert systems with applications*, *184*, 115537. https://doi.org/10.1016/j.eswa.2021.115537

[10] Hu, Z., Zhao, Y., & Khushi, M. (2021). A survey of forex and stock price prediction using deep learning. *Applied system innovation*, *4*(1), 1–30. https://doi.org/10.3390/asi4010009

[11] Fitriyanto, N., & Kusrini, K. (2025). Optimasi prediksi harga emas menggunakan CNN-Bi-LSTM dengan mekanisme attention dan bayesian optimization. *Journal of economic, management, accounting and technology*, *8*(1), 210–219. https://doi.org/10.32500/jematech.v8i1.8668

[12] Biswas, M., Shome, A., Islam, M. A., Nova, A. J., & Ahmed, S. (2021). Predicting stock market price: a logical strategy using deep learning. *2021 IEEE 11th IEEE symposium on computer applications & industrial electronics (ISCAIE)* (pp. 218–223). IEEE. https://doi.org/10.1109/ISCAIE51753.2021.9431817

[13] Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018). A comparison of arima and lstm in forecasting time series. *2018 17th IEEE international conference on machine learning and applications (ICMLA)* (pp. 1394–1401). IEEE. https://doi.org/10.1109/ICMLA.2018.00227

[14] Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. *Expert systems with applications*, *36*(3), 5932–5941. https://doi.org/10.1016/j.eswa.2008.07.006

[15] Saud, A. S., & Shakya, S. (2020). Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE. *Procedia computer science*, *167*, 788–798. https://doi.org/10.1016/j.procs.2020.03.419

[16] Esmaili, Z., & Torabipour, T. (2021). Providing a solution for predicting stock prices using lstm neural networks. *5th national conference on technology in electrical and computer engineering (TEC 2021)*. Civilica. https://civilica.com/doc/1281535

[17] Darapaneni, N., Paduri, A. R., Sharma, H., Manjrekar, M., Hindlekar, N., Bhagat, P., … & Agarwal, Y. (2022). *Stock price prediction using sentiment analysis and deep learning for Indian markets*. ArXiv Preprint ArXiv:2204.05783. http://dx.doi.org/10.48550/arXiv.2204.05783

[18] Zhong, Z., Wu, D., & Mai, W. (2023). Stock prediction based on ARIMA model and GRU model. *Academic journal of computing & information science*, *6*(7), 114–123. https://francis-press.com/uploads/papers/OckW11Te0djE8HywqAEKr9zwq4KUu1vpVHKMwM5L.pdf

[19] Yavasani, R., & Wang, H. (2023). Comparative analysis of LSTM, GRU, and ARIMA Models for stock market price prediction. *Journal of student research*, *12*(4), 1–9.

[20] Basher, S. A., & Sadorsky, P. (2022). Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility? *Machine learning with applications*, *9*, 100355. https://doi.org/10.1016/j.mlwa.2022.100351

[21] Ghaderzadeh, H., & Ghasiri Darbande, S. (2020). Investigating the effect of oil price fluctuations on agricultural products using the Arimax model and neural network during the period. *Agricultural economics research*, *12*(4), 189-204. (**In Persian**). https://www.ensani.ir/fa/article/432206/

[22] Mohammadi Kabari, S. M. H., Khosravi, A., & Dohli, S. (2024). Daily stock price prediction using improved lstm. *The first international conference on information technology, management and computers* (pp. 1-8). (**In Persian**). Civilica. https://civilica.com/doc/2084244