



COMPARATIVE ANALYSIS OF LSTM AND ARIMA MODELS IN STOCK PRICE PREDICTION: A TECHNOLOGY COMPANY EXAMPLE

Yasin KIRELLİ^{1*}


¹Kutahya Dumlupınar University, Management Information Systems, 43020, Kütahya, Türkiye

Abstract: Stock price forecasting has been an important area of interest for economists and computer scientists. In addition to traditional statistical methods, advanced artificial intelligence techniques such as machine learning can stand out with their ability to process complex data sets and adapt to historical data. In recent years, hybrid models combining deep learning and time series methods have demonstrated superior performance in stock selection and portfolio optimization. This study comparatively analyses the performance of LSTM and ARIMA models in time series forecasting. In the study, the stock prices of Oracle company are predicted using two different models, LSTM and ARIMA. Model performance is evaluated using metrics like MSE, MAE, RMSE, and MAPE. Both models have been found to be successful in different metrics. The LSTM model has lower error values; meanwhile, the ARIMA model produced proportionally more accurate forecasts. The study concludes that given the potential offered by deep learning, models such as LSTM are essential for time series forecasting. The flexibility of deep learning allows the development of customized models for different data types and time series problems.

Keywords: LSTM, ARIMA, Time Series, Stock Price Forecasting

*Corresponding author: Kutahya Dumlupınar University, Management Information Systems, 43020, Kütahya, Türkiye

E mail: yasin.kirelli@dpu.edu.tr (Y. KIRELLİ)

Yasin KIRELLİ  <https://orcid.org/0000-0002-3605-8621>

Received: March 01, 2024

Accepted: July 30, 2024

Published: September 15, 2024

Cite as: Kirelli Y. 2024. Comparative analysis of LSTM and ARIMA models in stock price prediction: a technology company example. BSJ Eng Sci, 7(5): xx-xx.

1. Introduction

Stock market price prediction is a research topic not only for economists but also for computer scientists, which is why statistical methods and machine learning are used to build a prediction model. Considering 2019 data, the global stock market is known to be more than 85 trillion dollars. Investors can use simple statistical methods to predict price changes to increase their income, but these methods could be faster due to the high volume of transactions and rapid changes. Advanced artificial intelligence methods are often preferred and used (Gandhmal and Kumar, 2019; Shah et al., 2019; Kanavos et al., 2020; Nabipour et al., 2020). The ability of the methods to predict with a high degree of accuracy is an important constraint. Therefore, machine learning methods are widely preferred. The most frequently preferred methods are logistic regression, linear regression, support vector machines, and artificial neural networks. It provides the advantage of creating strong, stable models for multidimensional problems and adaptability to historical data (Lu et al., 2021). In general, the stock market has a structure that combines high risk and high return. In this regard, investors aim to achieve maximum gains through technical and fundamental analysis methods. Fundamental analysis generally derives insights by considering the balance sheet analysis of companies, competitive advantages, or macroeconomic

developments. In technical analysis, price inferences are carried out with the assistance of indicators such as Moving Average Convergence Divergence (MACD), Momentum, Stochastic Oscillator with Moving Averages (KD), Relative Strength Index (RSI), and Moving Average (MA) (Rouf et al., 2021). Stock exchanges, where financial transactions take place, are regulated by a set of rules of governments and institutions to protect investors and companies. Stock prices operate according to the principles of supply and demand, and performance indicators (KPIs) depend on the rise and fall of prices, as shown in Table 1 (Bustos and Pomares-Quimbaya, 2020; Nti et al., 2020; Sarvesh et al., 2021).

Forecasting models for stock prices have attracted increased attention in recent years. Due to the financial opportunity provided, several papers have recently been published that use deep learning and time series methods to produce robust forecasting results. With the increasing power of today's machines and the improvement of computational capability, better results are being achieved for challenging problems of such complexity (Agrawal et al., 2013; Strader et al., 2020; Kumbure et al., 2022; Sheng et al., 2022). Singh et al. (2023) proposes a hybrid CNN-LSTM model combining deep learning for stock selection and the mean-variance model for portfolio optimization. Benchmarks on Indian stock market analysis show that the model outperforms single models



and other compared strategies. The results suggest that the hybrid CNN-LSTM approach can be an effective tool in portfolio management (P. Singh et al., 2023). Choi et al. (2023) propose a new method called "Hybrid Information Mixing Module," using deep learning and news data to predict stock price movements. It aims to make more accurate predictions by combining price and text data features and modeling their multimodal interactions. The model is tested on a high volatility stock market simulation, successfully validated, and performs better than other methods (Choi et al., 2023). Jadhav et al. (2021) proposes a new model using GAN architecture to predict stock market prices. The model consists of a hybrid generator and MLP parser using Naive Bayes for sentiment analysis from financial news and LSTM for prediction. The system predicts stock values using financial data and performs sentiment analysis on financial news (Jadhav et al., 2021). Singh et al. (2022) proposed a new investment decision support system by comparing ARIMAX, Prophet, LSTM, and pairwise LSTM models using historical stock market data. The model aims to maximize profit and minimize loss by optimizing investment decisions during crises such as COVID-19 with 98.60% and 96.97% accuracy (Singh et al., 2022). Choy et al. (2021) compared various time series algorithms to predict stock market prices: ARIMA, LSTM, SARIMA, Holt Winter, and Prophet. The research has been conducted on Malaysia stock market data covering six different sectors. The results show that the ARIMA and LSTM models achieve lower error rates than the others. Moreover, ARIMA model is more successful than LSTM especially in trend forecasting. This study suggests that the ARIMA model may be suitable for short-term stock price forecasting (Choy et al., 2021; Singh et al., 2022). Sisodia et al. (2022) propose a Deep Learning (DL) based Long Short-Term Memory (LSTM) algorithm to predict stock market price movements. The research develops a model

using ten years of historical price information of the NIFTY 50 index from the National Stock Exchange of India (NSE) data. After normalization, the training and testing process is performed, and an accuracy rate of 83.88% is obtained. These results show that the proposed model has the potential to predict stock market price movements (Sisodia et al., 2022). Ruan et al. (2020) propose a new method to predict sudden stock price changes, especially during unexpected events such as COVID-19. The study uses ARIMA and LSTM models to forecast individual stocks and the entire market. In addition, social media data reflecting investor psychology are also integrated into the models. The 98% average forecast accuracy obtained using the 2016-2020 period data of 100 companies in the study provides promising results for improving existing forecasting methods (Ruan et al., 2020).

Table 1. Key performance indicators

Indicator Name	Indicator Description
Open	Opening price
Close	End-of-day closing price
Low	Intraday low price
High	Intraday high price
Volume	Total number of stocks traded during the trading session

Using the Web of Science (WOS) database, the methods that have been used in stock market forecasting searched between 2012 and 2024, and the word cloud is shown in Figure 1.a. and Figure 1.b. obtained. LSTM and ARIMA models, which are also the subject of this study, have been explicitly analyzed, and it is clear from Figure 1.a. and Figure 1.b that the LSTM deep learning approach is the subject of more research area than the ARIMA model.

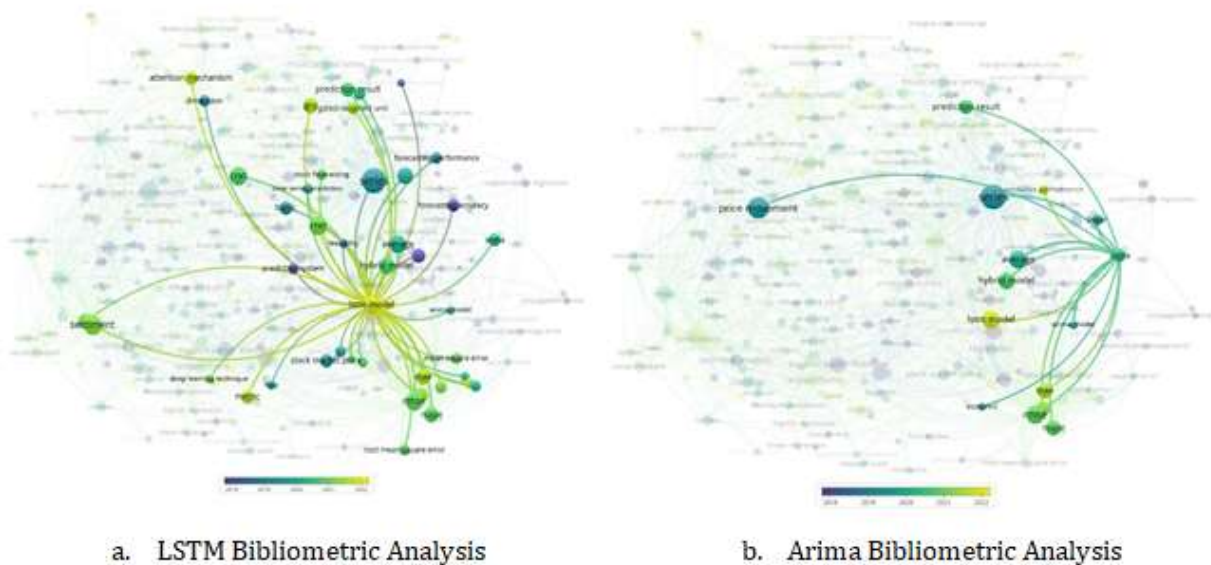


Figure 1. Bibliometric analysis of stock market forecasting methods.

2. Materials and Methods

This study constructs two models to forecast Oracle stock prices: LSTM and ARIMA. The objective is to evaluate these models' performance and ease of use and provide insights for financial analysts and investors. The background and basic principles of the LSTM and ARIMA models are explained in the first stage of the research. Then, both models are applied to daily data of Oracle stock prices, and their performance is compared using metrics such as MSE, MAE, RMSE, and MAPE. The results reveal the strengths and weaknesses of each model and demonstrate their success in predicting stock prices.

2.1. Long Short Term Memory (LSTM)

Long Short Term Memory (LSTM) networks are a specialized deep learning method, a Recurrent Neural Networks (RNN) sub-branch. It can store long-term dependencies and can maintain dependencies. Unlike RNNs, LSTM networks also store long-term dependencies, making them better suited to complex problems such as time series (Lipton et al., 2015; Sherstinsky, 2020; Van Houdt et al., 2020; Yang et al., 2020). In LSTM architecture, information is stored in cells, and the data's meaning can be interpreted according to the condition. An advantage of LSTMs is that they have a forget gate mechanism. Thus, it also solves the problem of gradient loss (Staudemeyer and Morris, 2019). This structure provides the advantage of giving successful results in time series or lengthy texts. As shown in Figure 2, an LSTM network has three main structures: the forget gate, the input gate, and the output gate.

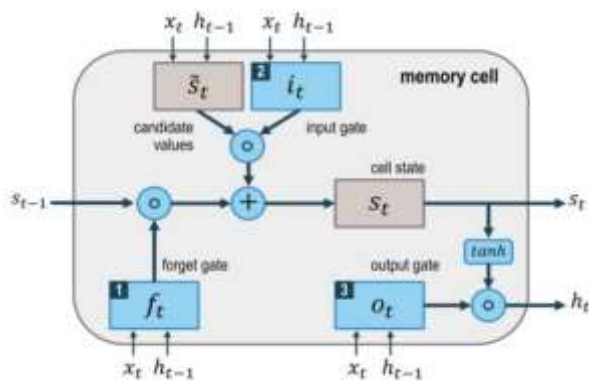


Figure 2. LSTM network.

The cell state stores past information and uses it for future prediction. The forget gate determines which information from the previous cell will be forgotten. The input gate decides which information is sent to the 'cell state'. The

output gate decides which information from the 'cell state' will be output. According to equations 1-5, x_t is the input vector, h_t is the output vector, f_t is the forget vector, and c_t is the cell state vector. i_t and o_t are the input and output gates, respectively. W and b are the weights and bias values (Ojo et al., 2019).

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2}$$

$$c_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{3}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{4}$$

$$h_t = \sigma * \tanh(c_t) \tag{5}$$

2.2. Autocorrelated Integrated Moving Average (ARIMA)

Autocorrelated Integrated Moving Average (ARIMA) is a statistical model. It is used as a method for analyzing and forecasting time series. The ARIMA model consists of autocorrelation (AR), integration (I), and moving average (MA) (Khan and Alghulaiakh, 2020). Autocorrelation is the relationship between past and present data in a time series, and its effect is measured. Integration refers to stationarity in a time series. It is provided by taking the difference from the time series' previous values. Moving average refers to random variation in a time series. This method converts time series to stationary through differencing and is called a linear stochastic model. This is generally expressed as ARIMA(p,d,q) (Kontopoulou et al., 2023; Kumar and Vanajakshi, 2015). In the model, d refers to the degree of differencing, and p and q refer to the autoregressive and moving average, respectively.

2.3. Proposed Models and Implementations

2.3.1. Exploratory data analysis

In the study, Oracle Company, which is listed on the Nasdaq Stock Exchange, is analyzed. The Nasdaq Stock Exchange is a New York-based American Stock Exchange and is among the most active in volume. It is traded as the second most active area after the New York Stock Exchange (Oracle Corporation (ORCL), 2024). The data set in the study has been obtained from the platform where the current trading prices of the companies listed on the Nasdaq official website are also available. The data set includes the opening and closing price, the highest and lowest price, and volume information of the symbol on a historical basis. The day-based closing price, which users mostly prefer, is evaluated in the training set. Table 2 shows a view of the data set derived for the prediction model.

Table 2. Dataset sample view

Date	Close/Last	Volume	Open	High	Low
01/29/2024	\$113.75	7014424	\$114.19	\$114.59	\$113.01
01/26/2024	\$114.64	5541900	\$114.64	\$115.48	\$114.2746
01/25/2024	\$115.00	8877205	\$114.89	\$116.18	\$114.35
01/24/2024	\$114.31	11702700	\$113.42	\$115.42	\$113.33
01/23/2024	\$111.83	9185921	\$110.29	\$112.68	\$109.56

2.3.2. LSTM model implementation

The analysis has been performed with a total of 2516 sample data for ten years between February 2014 and February 2024 of the company considered in the study. As shown in Figure 3, the company's time series is visualized according to the closing price and date values. The last year's data (after 2023) is divided into test data (10%), and up until 2023, it is divided into training data. In the first part of the study, the proposed LSTM and ARIMA models are used for forecasting in the second part. The last section gives the forecasting performance comparisons of the two models. This study evaluates the prediction results with the proposed LSTM model; Python

programming language version 3.10 is preferred. The model has been designed with the TensorFlow deep learning library. According to the proposed model, the training and test data have been normalized with the Min-Max scaler method in the pre-processing step, and accordingly, the data have been scaled between 0-1. The model and hyperparameters are shown in Table 3. The loss value of the model result is close to zero, and the validation loss value range is observed in the range of 0 - 0.002. The time series graph comparing the prediction values generated according to the proposed model and the actual values of the test set is shown in Figure 4.



Figure 3. Time series by training and test data.

Table 3. Model and hyperparameter values

Hyperparametres	Values	Model		
Scaling	Min-Max Scaler	Model: "sequential"		
Optimiser	Adam	Layer (type)	Output Shape	Param
Epoch	100	lstm (LSTM)	(None, 100, 50)	10400
Batch Size	64	lstm_1 (LSTM)	(None, 100, 50)	20200
		lstm_2 (LSTM)	(None, 50)	20200
		dense (Dense)	(None, 1)	51
		Total params: 50,851		
		Trainable params: 50,851		



Figure 4. LSTM Model test set and prediction values comparison time series plot.

2.3.3. ARIMA model implementation

As the second method created in the study, the ARIMA model aims to observe the forecasting performance by dividing the data set into training and test data sets at the same rate. According to the Akaike Information Criterion (AIC) criterion, in the stepwise search for the best ARIMA model, the lowest AIC value has been obtained with the ARIMA (0, 1, 1) model, as shown in Table 4. According to this model, past values ($p=0$) do not directly affect current values. It is assumed that the data are initially non-stationary ($d=1$) and have a moving average ($q=1$).

According to Figure 5, the time series is distributed between -8 and 11. Most of the residuals are concentrated around 0, suggesting that the model consistently fits the data. The residuals appear randomly distributed on the horizontal axis, concluding that the model has no autoregressive error. It is observed that the ARIMA (0, 1, 1) model is generally compatible with the time series data. As a result of the model training, the time series and forecast region evaluation graph of the 277-day test data from 2023 is observed, as shown in Figure 6.

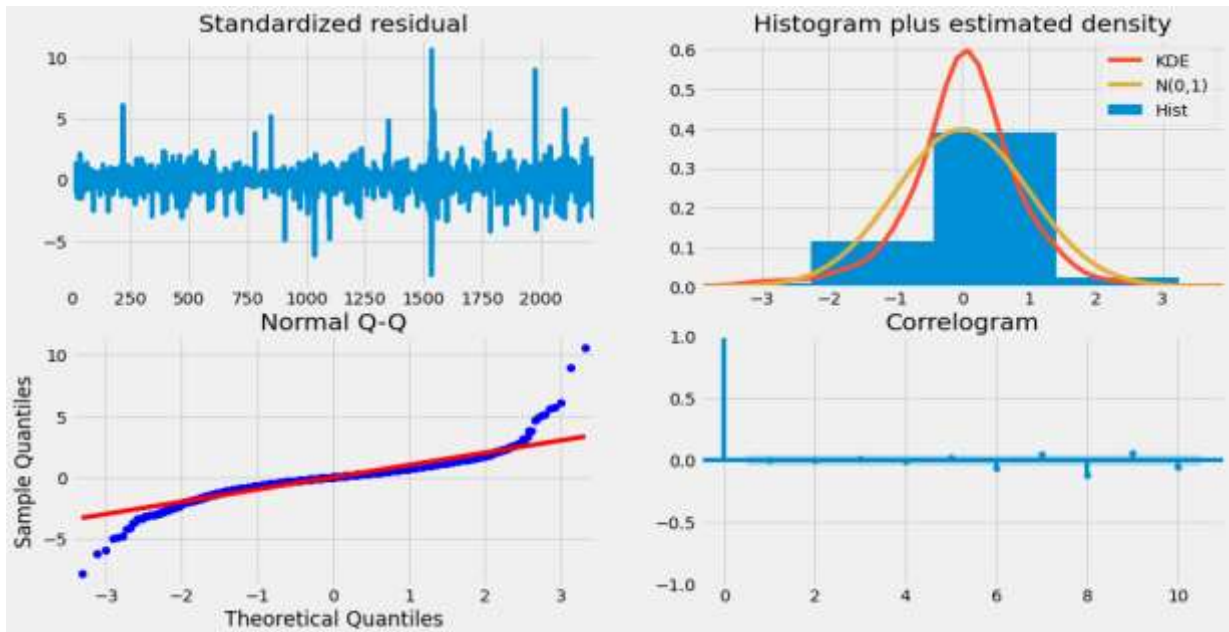


Figure 5. Standardised residual graph of data.

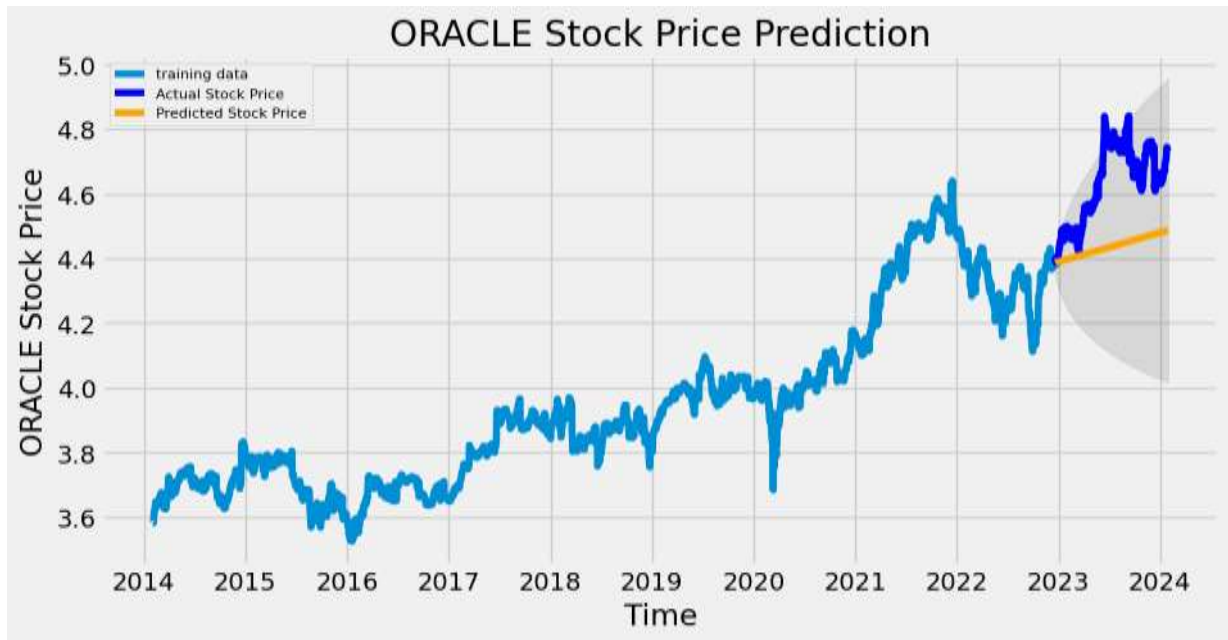


Figure 6. ARIMA model prediction.

Table 4. Stepwise search to minimize AIC

ARIMA(0,1,0)	AIC = -12070.156
ARIMA(1,1,0)	AIC = -12099.264
ARIMA(0,1,1)	AIC = -12099.390
ARIMA(0,1,0)	AIC = -12071.055
ARIMA(1,1,1)	AIC = -12097.461
ARIMA(0,1,2)	AIC = -12097.461
ARIMA(1,1,2)	AIC = -12095.459
ARIMA(0,1,1)	AIC = -12099.961
ARIMA(1,1,1)	AIC = -12098.046
ARIMA(0,1,2)	AIC = -12098.049
ARIMA(1,1,0)	AIC = -12099.874
ARIMA(1,1,2)	AIC = -12096.046

3. Results and Discussion

Different metrics can be used to evaluate the performance of time series forecasting models, the most common of which are:

Mean Squared Error (MSE): Expressed as the mean squared error, it is the average sum of squares between the predicted (\hat{y}) and actual (y_i) values as equation 6.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2 \quad (6)$$

Mean Absolute Error (MAE): Refers to the mean of the absolute value between the predicted actual values as equation 7.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}| \quad (7)$$

Root Mean Squared Error (RMSE): The square root of the root mean squared error is a better representation of the scale of the amount of error as equation 8.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2} \quad (8)$$

Mean Absolute Percentage Error (MAPE) Refers to the mean absolute percentage between the predicted and actual values as equation 9.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \quad (9)$$

The comparison of metric values for the two models included in the study is observed in Table 5.

Table 5. LSTM and ARIMA model performance metrics

	MSE	MAE	RMSE	MAPE
LSTM Model	0.0374	0.1700	0.1936	0.1855
ARIMA Model	0.0488	0.1968	0.2211	0.0419

According to the comparative result metrics, both models have low MSE values, which supports the consistency of the models. The LSTM model has a lower MAE value than the ARIMA model, which means that the LSTM model

makes less error on average. The results show that the LSTM model tends to produce estimates with less error and variation than the ARIMA model. The study compares the performance of LSTM and ARIMA models in time

series forecasting. The findings show that both models stand out in specific metrics. The choice of model depends on the prioritized metrics and the dataset's characteristics. The LSTM model generally produces fewer error and variation forecasts than the ARIMA model. In contrast, the ARIMA model tends to produce proportionally more accurate forecasts. The results of this study have important implications for investors. Deep learning models such as LSTM, improve decision-making processes for investors by reducing errors in price prediction. Such strategic approaches are also an indicator for ensuring risk management. Advanced forecasting models also provide portfolio optimization, conscious decision-making capabilities and competitive advantage for investors. Given the potential offered by deep learning, models such as the LSTM are essential for future time series forecasting. Future research could focus on comparing the performance of LSTM and ARIMA models on different data sets and combining them into a hybrid model. Developing advanced models for time series forecasting is also an essential area of research. Consequently, it is inferred that both LSTM and ARIMA models are practical tools for time series forecasting based on the literature. The flexibility of deep learning allows the development of customized models for different data types and time series problems.

Author Contributions

The percentage of the author(s) contributions is presented below. The author reviewed and approved the final version of the manuscript.

	Y.K.
C	100
D	100
S	100
DCP	100
DAI	100
L	100
W	100
CR	100
SR	100
PM	100

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management.

Conflict of Interest

The author declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

References

Agrawal JG, Chourasia V, Mittra A. 2013. State-of-the-art in stock prediction techniques. *Int J Adv Res Electr Electron Instrum Eng*, 2(4): 1360-1366.

Bustos O, Quimbaya A. 2020. Stock market movement forecast: A systematic review. *Expert Syst Appl*, 156: 113464.

Choi J, Yoo S, Zhou X, Kim Y. 2023. Hybrid information mixing module for stock movement prediction. *IEEE Access*, 11: 28781-28790.

Choy YT, Hoo MH, Khor KC. 2021. Price prediction using time-series algorithms for stocks listed on Bursa Malaysia. In: 2nd International Conference on Artificial Intelligence and Data Sciences, 8-9 September, Piscataway, New Jersey, USA, pp: 1-5.

Gandhmal DP, Kumar K. 2019. Systematic analysis and review of stock market prediction techniques. *Comput Sci Rev*, 34: 100190.

Jadhav R, Sinha S, Wattamwar S, Kosamkar P. 2021. Leveraging Market Sentiment for Stock Price Prediction using GAN. In: 2nd Global Conference for Advancement in Technology, 01-03 October, 2021, Bangalore, India, pp: 1-6.

Kanavos A, Vonitsanos G, Mohasseb A, Mylonas P. 2020. An entropy-based evaluation for sentiment analysis of stock market prices using Twitter data. In: 15th International Workshop on Semantic and Social Media Adaptation and Personalization, 29-30 October Zakynthos, Greece, pp: 1-7.

Khan S, Alghulaiakh H. 2020. ARIMA model for accurate time series stocks forecasting. *Int J Adv Comput Sci Appl*, 11(7).

Kontopoulou V, Panagopoulos AD, Kakkos I, Matsopoulos GK. 2023. A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. *Future Int*, 15(8): 255.

Kumar SV, Vanajakshi L. 2015. Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *Eur Transp Res Rev*, 7: 1-9.

Kumbure MM, Lohrmann C, Luukka P, Porras J. 2022. Machine learning techniques and data for stock market forecasting: A literature review. *Expert Syst Appl*, 197: 116659.

Lipton ZC, Kale DC, Elkan C, Wetzel R. 2015. Learning to diagnose with LSTM recurrent neural networks. *ArXiv*, 2015: 1511.

Lu W, Ge W, Li R, Yang L. 2021. A Comparative Study on the Individual Stock Price Prediction with the Application of Neural Network Models. In: International Conference on Computer Engineering and Artificial Intelligence, 27-29 August, Shanghai, China, pp: 235-238.

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

Nti IK, Adekoya AF, Weyori BA. 2020. A systematic review of fundamental and technical analysis of stock market predictions. *Artif Intell Rev*, 53(4): 3007-3057.

Ojo SO, Owolawi PA, Mphahlele M, Adisa JA. 2019. Stock market behaviour prediction using stacked LSTM networks. In: International Multidisciplinary Information Technology and Engineering Conference, 21-22 November, Vanderbijlpark, South Africa, pp: 1-5.

Oracle Corporation Stock Historical Prices & Data Yahoo Finance. URL: <https://finance.yahoo.com/quote/ORCL/history> (accessed date: January 29, 2024).

Rouf N, Malik MB, Arif T, Sharma S, Singh S, Aich S, Kim HC. 2021. Stock market prediction using machine learning techniques: a decade survey on methodologies, recent developments, and future directions. *Electronics*, 10(21): 2717.

Ruan J, Wu W, Luo J. 2020. Stock Price Prediction Under

- Anomalous Circumstances. In: IEEE International Conference on Big Data, pp: 4787-4794.
- Sarvesh S, Sidharth RV, Vaishnav V, Thangakumar J, Sathyalakshmi S. 2021. A hybrid model for stock price prediction using machine learning techniques with CNN. In: 5th International Conference on Information Systems and Computer Networks, pp: 1-6.
- Shah D, Isah H, Zulkernine F. 2019. Stock market analysis: A review and taxonomy of prediction techniques. *Int J Financ Stud*, 7(2): 26.
- Sheng Y, Fu K, Wang L. 2022. A PCA-LSTM Model for Stock Index Forecasting: A Case Study in Shanghai Composite Index. In: 7th International Conference on Cloud Computing and Big Data Analytics, pp: 412-417.
- Sherstinsky A. 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D*, 404: 132306.
- Singh A, Bansal A, Nair A, Kaushal A. 2022. Predictive Analytics of Stock Market as a Time Series. In: 4th International Conference on Advances in Computing, Communication Control and Networking, pp: 770-777.
- Singh P, Jha M, Sharaf M, El-Meligy MA, Gadekallu TR. 2023. Harnessing a Hybrid CNN-LSTM Model for Portfolio Performance: A Case Study on Stock Selection and Optimization. *IEEE Access*, 11: 104000-104015.
- Sisodia PS, Gupta A, Kumar Y, Ameta GK. 2022. Stock market analysis and prediction for NIFTY50 using LSTM Deep Learning Approach. In: 2nd international conference on innovative practices in technology and management, pp: 156-161.
- Staudemeyer RC, Morris ER. 2019. Understanding LSTM a tutorial into long short-term memory recurrent neural networks, *ArXiv*, 2019: 1909.
- Strader TJ, Rozycki JJ, Root TH, Huang YHJ. 2020. Machine learning stock market prediction studies: review and research directions. *J Int Technol Inf Manage*, 28(4): 63-83.
- Sanboon T, Keatruangkamala K, Jaiyen S. 2019. A deep learning model for predicting buy and sell recommendations in stock exchange of thailand using long short-term memory. In: IEEE 4th International Conference on Computer and Communication Systems, pp: 757-760.
- Van Houdt G, Mosquera C, Nápoles G. 2020. A review on the long short-term memory model. *Artif Intell Rev*, 53(8): 5929-5955.
- Yang S, Yu X, Zhou Y. 2020. Lstm and gru neural network performance comparison study: Taking yelp review dataset as an example. In: International workshop on electronic communication and artificial intelligence, pp: 98-101.