Cloud Computing and Data Science
http://ojs.wiserpub.com/index.php/CCDS/

UNIVERSAL WISER PUBLISHER

Research Article

Advancing Stock Market Predictions with Time Series Analysis Including LSTM and ARIMA

Ishtiaq Ahammad[1*], William Ankan Sarkar[2], Famme Akter Meem[2], Jannatul Ferdus[2], Md. Kawsar Ahmed[2], Md. R. Rahman[2], Rabeya Sultana[2], Md. Shihabul Islam[2]

[1]Department of Computer Science and Engineering, Northern University Bangladesh, Dhaka, Bangladesh
[2]Department of Computer Science and Engineering, Prime University, Dhaka, Bangladesh
E-mail: ahammadishtiaq27@gmail.com

Received: 19 February 2024; **Revised:** 22 April 2024; **Accepted:** 24 April 2024

Abstract: Predicting stock market prices accurately is a major task for investors and traders seeking to optimize their decision-making processes. This research focuses on the comparative analysis of advanced machine learning (ML) techniques, particularly, the Long Short-Term Memory (LSTM) model and Autoregressive Integrated Moving Average (ARIMA) model for predicting stock market prices. The study enforces thorough data collection and preprocessing to ensure the quality and reliability of the historical stock price data, forming a robust foundation for the predictive models. The core contribution of this paper lies in its systematic and comparative analysis of these two models. A range of performance metrics, including Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), are employed to assess and contrast the predictive accuracy and efficiency of the LSTM and ARIMA models. The research findings indicate that the ARIMA model, contrary to expectations, outperforms the LSTM model in this study, achieving lower RMSE and MAE values. Specifically, the ARIMA model demonstrates a Test RMSE of 4.336 and a Test MAE of 3.45926, indicating its superior predictive accuracy compared to the LSTM model. Furthermore, the study sets its findings against the backdrop of existing literature by comparing the performance of its models with those reported in previous research. This comparison shows better results achieved by our stock market prediction models. By addressing limitations observed in prior studies and demonstrating practical applicability, this research contributes to advancing stock market prediction methodologies, offering valuable insights for investors and traders.

***Keywords*:** stock market prediction, time series analysis, LSTM, ARIMA, machine learning, predictive modeling, comparative analysis

1. Introduction

In the context of the stock market, where the environment is characterized by volatility and complexity, investors are constantly striving to optimize their profits through informed decision-making. Predicting future stock price movements has always posed a formidable challenge due to the intricate interplay of economic, political, and social factors [1]. However, advancements in time series analysis and machine learning (ML) techniques have paved the way for the development of more sophisticated and accurate prediction models [2].

Stock market prediction, commonly referred to as forecasting, involves the attempt to anticipate the future value of

DOI: https://doi.org/10.37256/ccds.5220244470

a business's stock or any other financial instrument traded on an exchange [3]. Accurate predictions of a stock's future price can potentially yield substantial profits [4]. ML, a subset of artificial intelligence (AI), leverages vast datasets from the past, uncovers latent relationships within them, forecasts future trends, and continually refines its predictive capabilities [5]. ML techniques have applications in vast domains including education [6-9], medical [10-11], sports [12], manufacturing [13], and stock market [14]. In the realm of stock trading, ML has gained increasing prominence, with investment firms utilizing it for a large number of purposes, including predicting market fluctuations, analyzing customer behavior, and scrutinizing stock price dynamics [15]. Notably, one of the most critical applications of ML is in the domain of stock price analysis [16]. Nevertheless, predicting stock prices using ML is a formidable task, given the countless physical, psychological, and other dynamic factors that influence stock values. Consequently, generating precise stock price forecasts remains a challenging endeavor [17-18].

1.1 *Motivation of the research*

This research is motivated by the ongoing difficulties in predicting the stock market, which are made tougher by the unpredictable and dynamic nature of financial markets. Because of the complex interactions between political, social, and economic issues, reliable stock price forecasting is still difficult to achieve, even with advances in time series analysis and ML approaches [19]. A number of shortcomings have been noted by earlier research, such as the absence of thorough model comparisons, poor performance scores, a lack of features and data sources, insufficient comparison analysis with earlier studies, and a lack of practical application. Therefore, it is imperative to address these issues and progress the field of stock market forecasting by creating more reliable and accurate forecasting models. This research aims to address these challenges by conducting a thorough comparison between LSTM and ARIMA models, leveraging historical stock market data for a comprehensive evaluation, delivering comparative analyses with relevant previous studies, and demonstrating the practical applicability of the proposed models in real-world trading scenarios.

1.2 *Contributions of the research*

Various ML models have been employed for stock market predictions, each carrying its own set of advantages and limitations. The details of these previous works as well as their limitations and research gaps are discussed in Section 2. In that section, we address the gaps identified in prior stock market prediction research and outline how our study addresses them. Many previous studies lack comprehensive model comparisons, focusing instead on individual techniques, leaving uncertainty about optimal model performance. Our research fills this void by thoroughly comparing Long Short-Term Memory (LSTM) and Autoregressive Integrated Moving Average (ARIMA) models, shedding light on their respective strengths and weaknesses. Moreover, prior studies often exhibit poor performance scores with evaluation metrics like MAE and RMSE, hindering accurate prediction assessment. In contrast, our study demonstrates significantly improved performance in terms of both MAE and RMSE, offering more reliable predictions. Additionally, while previous works may overlook valuable data sources and features, our study utilizes historical stock market data to provide a more comprehensive evaluation of model accuracy. Furthermore, we conduct comparative analyses with relevant previous studies, ensuring a clearer understanding of the superior performance of our LSTM and ARIMA models. Lastly, our research addresses the lack of real-world applicability observed in many previous studies by demonstrating the practical effectiveness of our models in predicting stock market prices, thus offering valuable insights for investors and traders. Overall, our study contributes to advancing stock market prediction methodologies by addressing these existing limitations and providing a more robust framework for analysis.

1.3 *Brief discussion on research procedure and outcome*

The initial step of our research's anticipated strategy for building the prediction system is data collection and processing. We gathered data for the past ten years from the Apex Foods firm. The information contains date, last_ traded_price, high, low, opening_price, closing_price, yesterdays_closing_price, trade, value_mn, volume. The term "high and low" refers to a share's highest and lowest value for a given day. After that, the preparation of data is a crucial stage that needs to be carried out carefully and correctly. All of the collected raw datasets are initially contained in the Excel file. The data is then cleaned in accordance with our requirements. Then we apply our time series techniques to our collected and processed data. On the premise that future trends will be similar to historical trends, time series techniques make predictions about the future by examining historical trends. In our research, the considered models are LSTM and ARIMA. When predicting future stock prices using LSTM and ARIMA, past data is used to train and test the models. We have divided datasets into a training set (70%) and a test set (30%) for stock market prediction. The prediction performance of the LSTM or ARIMA model is tested using testing data after training. The model has not seen the testing data because it is a different collection of historical stock market data. It offers information on the trained model's generalization to new data as well as its capacity to predict future stock values. In order to ensure consistency in the input format, the testing data is pre-processed in the same manner as the training data, including scaling for LSTM or differencing for ARIMA. The model is then used to predict future stock prices using the testing data. For assessing the performance of our models, we have utilized Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). These measures aid in comparing the efficacy of our models and offer insights into prediction mistakes.

The goal of our study is to predict closing stock market values. Our research presents a comprehensive comparison of the performance of LSTM and ARIMA models in predicting stock market prices. The effectiveness of the models was evaluated using MAE and RMSE as the primary metrics. For the LSTM model, the results demonstrated a Test RMSE of 141.45199017262152 and a Test MAE of 139.95950777559275. On the other hand, the ARIMA model showed superior performance with significantly lower error rates, achieving a Test RMSE of 4.336 and a Test MAE of 3.45926. The stark contrast in the error metrics underscores the more precise predictions made by the ARIMA model in comparison to the LSTM model. In addition, a comparison of our research's outcome with other relevant works' outcomes are also performed. This comparative analysis highlights the superior performance of our research's LSTM and ARIMA models, indicating a significant advancement in accuracy for stock price prediction compared to previous works.

1.4 *Structure of the research*

The structure of this research paper is as follows: The research's problem definition, motivation, and novelty are presented in section 1. Current research works along with their limitations in the area of stock market prediction using ML approaches and time series analysis are reviewed in section 2. The prerequisites and procedure for implementing and simulating our research are provided in section 3. The results and discussions of the applied ML models are presented in section 4. Finally, section 5 presents the conclusion and future research scope.

2. Literature review

In this literature review, we examine studies relevant to the scope of our research. In order to have a better understanding of a particular research subject, we must focus on the most recent literature that is published. Therefore, as we try to understand the details of papers related to stock market forecasting assessments, the majority of our focus has been on the most recent works in the domain. After performing the recent relevant works, we present the research gaps of these studies as well as how these gaps are overcome by our study.

2.1 *Related works*

The study [20] explores a method for stock market prediction by employing LSTM-RNN models and ensemble techniques. It assesses various data sources and applies Differential Evolution and Weighted Average methods for predicting stock prices accurately over different time horizons, ranging from one to thirty days. Additionally, the research combines deep learning (DL) and ML approaches to create hybrid models for stock price prediction. Notably, the most accurate model utilizes LSTM networks and data from the previous week to forecast the open value of the NIFTY 50 index. In the research presented in [21], a hybrid modeling approach combining DL and ML techniques was employed for the prediction of stock prices. The specific focus of this research was on forecasting the open values of the NIFTY 50 index records. To achieve this, eight regression models were utilized in the analysis. Notably, four of these regression models were constructed using DL methodologies, specifically LSTM networks. To evaluate the performance of these models, a unique walk-forward method was applied. The key outcome of this investigation revealed that the most accurate predictive model was an LSTM-based univariate model. The research [22] adopts a comprehensive approach, integrating DL, ML, and statistical models. It evaluates eight ML and statistically-based models, of which, Convolutional neural network (CNN) models outperform LSTM in terms of accuracy.

In [23] authors introduce data mining (DM) techniques, including multiple regression, polynomial regression, linear regression, and prediction, as tools to uncover hidden patterns and enhance accuracy in stock market prediction. The study explores regression techniques such as Regression, Polynomial Regression, and RBF Regression. It highlights the value of DM in scenarios where conventional and statistical methods may fall short. Another similar work that focuses on DM is performed at [24]. In this study, DM techniques are combined with ML and AI to predict stock price movements. Sentiment analysis of tweets received via the Twitter API and the closing prices of equities are used to forecast stock prices. Another work that focuses on social media data as input is performed at [25]. By investigating the impact of social media and financial news data on stock market predictions, this study employs algorithms and feature selection techniques. The results indicate that financial news and social media contribute significantly to forecast accuracy, with the random forest classifier demonstrating high reliability.

The research [26] employs Support Vector Machine (SVM) technology to estimate stock prices in various markets, including both large and small capitalizations. The objective is to reduce uncertainty in investment decisions. The study encompasses daily and up-to-the-minute price frequencies and demonstrates the efficiency of SVM in prediction, with no overfitting issues observed. Similar work which includes SVM is also presented by [27]. In this work, market performance is predicted based on factors such as interest rates, foreign exchange rates, oil prices, and social media data. ML approaches, including SVM, Radial Basis Function (RBF), Multi-Layer Perceptron (MLP), and Single Layer Perceptron (SLP), are utilized. MLP emerges as the top performer, accurately predicting 77% of market performance.

In [28], authors explore various classifiers using both normal and leaked datasets for stock market prediction. It delves into methods such as technical analysis, time-series forecasting, ML, DM, and stock volatility modeling. The research combines datasets from different stock markets with seven classifiers to anticipate market trends, followed by sentiment analysis to refine predictions. Sentiment analysis for stock price prediction was particularly focused on [29]. In this study, a multi-faceted approach was adopted for stock market prediction. Initially, various predictive approaches, including sentiment analysis, were employed to gather insights into market dynamics. Sentiment analysis was utilized to compute polarity scores, which served as valuable inputs for predicting stock prices. Notably, a Random Forest approach was chosen as the primary predictive model. This predictive journey commenced with elementary techniques such as averaging and linear regression, which laid the foundation for more sophisticated modeling approaches. Significantly, the study revealed that the Random Forest method surpassed logistic regression, particularly in the context of sentiment analysis for stock market prediction.

The study [30] focuses on predicting a stock's closing price for the next day using Random Forest and Artificial Neural Network (ANN) algorithms. It incorporates various stock price variables as inputs and assesses model performance using metrics like RMSE and MAPE. ANN is also used to anticipate the closing price of the stock the next day and is also utilized for comparative research [31]. The results (0.013) demonstrate that the optimal values produced by the ANN model are RMSE (0.42), MAPE (0.77), and MBE. Another research that explores stock market prediction methodologies, including ANN and NN procedures can be found at [32]. It emphasizes the use of hybrid approaches to improve prediction accuracy, highlighting the influence of government decisions and consumer opinions on stock markets.

2.2 *Research gap and this study's solution*

The limitations and research gaps of previous works on stock market prediction as well as how our study overcomes them are presented below.

Model Comparison: Most of the previous studies focus on individual models or techniques without conducting comprehensive comparisons with alternative approaches. This can lead to uncertainty about which model performs best under different market conditions. Conducting a thorough comparison between LSTM and ARIMA models, our study provides insights into the strengths and weaknesses of each approach.

Poor Performance Score: Previous studies that uses MAE and RMSE evaluation metrics as the performance parameter; deliver very poor performance scores. This can make it difficult to assess the robustness and accuracy of their predictions. However, our research provides better performance compared to previous studies in terms of MAE and RMSE. Specifically, our study demonstrates significantly lower error rates for both MAE and RMSE when comparing the LSTM and ARIMA models.

Data Sources and Features: The choice of data sources and features used in previous studies are limited, potentially overlooking valuable information that could improve prediction accuracy. In our study, employing historical stock market data to evaluate model accuracy, we ensure a comprehensive assessment of both LSTM and ARIMA models.

Comparative Analysis with Previous Works: The above studies don't deliver a proper comparison between their and previous works. Conducting a comparative analysis with other relevant studies to demonstrate superior performance is a vital step in prediction-related studies. In our study, we deliver comparative performance among LSTM and ARIMA models, as well as their scores with the relevant previous studies.

Lack of Real-world Application: Several previous works lack practical applicability and fail to demonstrate the effectiveness of their proposed methods in real-world trading scenarios. In our study, by demonstrating the practical applicability of LSTM and ARIMA models in predicting stock market prices, we offer valuable insights for investors and traders.

Overall, our research contributes to addressing existing limitations and advancing the field of stock market prediction by offering a comprehensive comparison of LSTM and ARIMA models and demonstrating their effectiveness in real-world scenarios.

3. Methodology

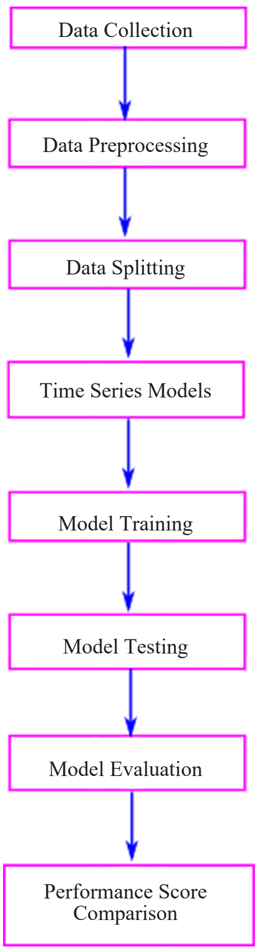

Figure 1. Workflow Diagram of this Research

The main content of the research we've conducted is demonstrated in this section. Along with other necessary discussions, the two key components of our research's anticipated strategy for building the prediction system are presented below. These two key components are Data collection and processing and Time series techniques used. After that, a discussion about the training and testing, data splitting, and utilized performance metrics are carried out. The workflow of this research is presented in Figure 1.

3.1 *Data collection*

Table 1. Sample Data

date	last_traded_price	high	low	opening_price	closing_price	yesterdays_closing_price	trade	value_mn	volume
12/5/20110:00	69.1	71.6	67.8	71.6	69.2	70.9	27	0.1876	2700
12/7/20110:00	69.4	71	62.7	64.1	69.4	69.2	43	0.3437	5000
12/8/20110:00	65.6	72	65.5	70	65.7	69.4	137	0.9577	14200
12/11/20110:00	65	66.5	63.7	63.7	64.8	65.7	41	0.5924	9150
12/12/20110:00	67	69.4	64.1	66.8	67.4	64.8	44	0.5174	7700
12/13/20110:00	71.4	72.9	69.5	69.9	71.5	67.4	197	1.5789	22200
12/14/20110:00	70.8	74	69	74	69.6	71.5	170	1.4234	20050
12/15/20110:00	68.5	71.5	66.6	69.1	68.3	69.6	52	0.4765	6950
12/18/20110:00	70.9	70.9	66.5	66.5	69.8	68.3	85	0.774	11250
12/19/20110:00	76.7	76.7	71.3	71.8	76.6	69.8	230	2.4828	33150
12/20/20110:00	81.3	83	73.7	78.1	80.8	76.6	215	2.1784	26950
12/21/20110:00	83.8	85.8	80	81.7	82.6	80.8	243	2.783	33400
12/22/20110:00	81.8	83.7	80.3	81.7	81.7	82.6	109	1.6809	20450
12/26/20110:00	80.1	83.2	80.1	83.2	80.7	81.7	183	1.239	15150
12/27/20110:00	82.7	82.8	80.5	80.5	82.4	80.7	139	1.4487	17700
12/28/20110:00	80	84	80	84	80.4	82.4	81	0.8315	10300
12/29/20110:00	82.6	84	80.2	81.1	83	80.4	226	1.9846	24050
1/1/20120:00	85.4	89.9	84	84	86.7	83	230	2.3653	27100
1/2/20120:00	91.5	91.5	86.8	86.8	89.9	86.7	281	3.1948	36050
1/3/20120:00	91.1	92	89	90	90.9	89.9	261	3.4911	38400
1/4/20120:00	89	91	88.7	91	89.7	90.9	157	1.9081	21300
1/5/20120:00	88.7	91	87	90	88.1	89.7	164	2.3445	26200
1/8/20120:00	87.1	90	87	89.6	87.7	88.1	97	1.3178	14800
1/9/20120:00	84.4	88.6	84.4	86.8	84.5	87.7	85	0.8007	9350
1/10/20120:00	81.6	86.1	80.3	85.1	82.1	84.5	99	1.2184	14700
1/11/20120:00	83.5	86	80.3	81.2	84.5	82.1	104	1.5423	18600

A trustworthy source with enough and pertinent data is crucial since the goal of this study is to anticipate the stock's value based on its historical value and trends. The Kaggle [33] is our primary source of data. The information contains date, last_traded_price, high, low, opening_price, closing_price, yesterdays_closing_price, trade, value_mn, volume. The term "high and low" refers to a share's highest and lowest value for a given day. We gathered data for the past ten years from the Apex Foods firm. The preparation of data is a crucial stage that needs to be carried out carefully and correctly. All of the collected raw datasets are initially contained in the Excel file. The data is then cleaned in accordance with the requirements. The data sample is shown below in Table 1. This sample data section includes a graphical table with essential columns that provide information about the financial market activities of our considered dataset. It contains data such as the date, last traded price, high and low, opening and closing prices, yesterday's closing price, trade, value_mn, and volume.

3.2 *Time series techniques*

Time series analysis is a vital statistical tool for understanding data patterns over time. Time series analysis provides valuable insights into temporal data, aiding decision-making processes in various fields like finance, economics, and environmental science. By examining sequential data points, such as stock prices or monthly sales figures, analysts can uncover trends, seasonality, and irregular fluctuations. This analysis often involves techniques like trend analysis, seasonal decomposition, and forecasting methods. Trend analysis helps identify long-term movements, while seasonal decomposition separates periodic patterns from the overall trend. Forecasting methods, such as LSTM, and ARIMA, enable predictions based on historical data. Since the foundation of our research is the prediction of the stock market, we must be aware of the time series techniques that are frequently employed in stock market prediction. The following is an example of many time series techniques that are frequently applied to stock market forecasting: Long Short-Term Memory (LSTM) Neural Networks, Recurrent Neural Networks (RNN), Random Forests (RF), Autoregressive Integrated Moving Average (ARIMA), Support Vector Machine (SVM), Moving Average, Vector Autoregression (VAR), Seasonal Autoregressive Integrated Moving Average (SARIMA) etc. These can be considered but a few illustrations of time series techniques might be applied to stock market forecasting. Our considered models are LSTM and ARIMA. These algorithms enable us to successfully forecast the stock market, which is our main objective.

3.2.1 *Long short-term memory (LSTM) model*

The LSTM model is a type of recurrent neural network (RNN) architecture designed to capture long-term dependencies in sequential data. This makes it particularly effective for time series analysis, natural language processing, and speech recognition. Unlike traditional RNNs, LSTMs have specialized memory cells that can store information over extended time periods, allowing them to remember important past events while selectively forgetting irrelevant information. This capability makes LSTMs well-suited for tasks where understanding context or remembering past interactions is crucial. By learning from sequential patterns in data, LSTMs can make accurate predictions and generate meaningful outputs, contributing significantly to advancements in various fields like the stock market.

3.2.2 *Autoregressive integrated moving average (ARIMA)*

The Autoregressive Integrated Moving Average (ARIMA) model is a widely used time series forecasting technique known for its effectiveness in capturing complex temporal patterns. It combines autoregression, differencing, and moving average components to model and predict time series data. ARIMA models are particularly suitable for stationary time series, where statistical properties like mean and variance remain constant over time. By analyzing past observations and identifying trends, seasonality, and irregularities, ARIMA models can make accurate predictions about future values. This versatility has made ARIMA a popular choice in various fields, including finance, economics, and environmental science, where understanding and forecasting sequential data is essential for decision-making. Specifically talking, it is useful for predicting stock prices that have a more complex pattern of behavior.

3.3 *Training and testing*

One serious fact about the stock market is that it has an unstable structure. This makes the prediction work much more difficult. When predicting future stock prices using LSTM and ARIMA, past data is used to train and test the models.

Historical information on stock prices and other pertinent variables is needed to train an LSTM or ARIMA model for stock market prediction. A time series dataset including information on previous stock prices, trading volumes, and any additional elements that could have an impact on the stock's behavior makes up the training data. Depending on the intended prediction horizon, the data is often gathered over a set time period that might range from several months to many years. To make sure that all variables are of a comparable size and can be successfully trained by the model, the training data for LSTM is preprocessed by scaling the input features. The data is then divided into input sequences and associated goal values, where the desired value is the stock price at a given future date and the input sequence is a window of historical data. The autoregressive (AR) and moving average (MA) components of the ARIMA model are all estimated using the training data. Data is differentiated to make it stationary, which is necessary for ARIMA modeling, and to eliminate any patterns or seasonality. The parameters of the model are then calculated using techniques such as least squares estimation or maximum likelihood estimation. We have divided datasets into a training set (70%) and a test set (30%) for stock market prediction.

The prediction performance of the LSTM or ARIMA model is tested using testing data after training. The model has not seen the testing data because it is a different collection of historical stock market data. It offers information on the trained model's generalization to new data as well as its capacity to predict future stock values. In order to ensure consistency in the input format, the testing data is preprocessed in the same manner as the training data, including scaling for LSTM or differencing for ARIMA. The model is then used to predict future stock prices using the testing data.

3.4 *Performance analysis metrices*

The accuracy and dependability of the LSTM or ARIMA model for stock market prediction may be evaluated by analysts and researchers by comparing the model's predictions against the actual stock prices in the testing data. The model's performance evaluation aids in assessing its usefulness and directs prospective enhancements or alterations to the model's parameters or design. Different performance measures, including mean absolute error (MAE), mean squared error (MSE), and root mean square error (RMSE), can be used to assess how well LSTM and ARIMA models predict stock market values. These measures aid in comparing the efficacy of various models and offer insights into prediction mistakes.

3.4.1 *Mean absolute error (MAE)*

A measure of errors between paired observations describing the same phenomena is called mean absolute error (MAE). The formula for MAE is:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |(Y_i - \widetilde{Y}_i)|}{n} \tag{1}$$

MAE is Mean Absolute Error
Y_i is the actual value of the target variable for the i^{th} observation
$\widetilde{Y}_i$ is the predicted value of the target variable for the i^{th} observation
n is the number of observations
$\sum$ represents the summation symbol

3.4.2 *Mean squared error (MSE)*

The average squared difference between the observed and predicted outcomes is used to calculate the mean squared error (MSE), which quantifies statistical model error. The formula for MSE is:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widetilde{Y}_i)^2 \quad (2)$$

MSE is Mean Squared Error
Y_i is the actual value of the target variable for the i^{th} observation
$\widetilde{Y}_i$ is the predicted value of the target variable for the i^{th} observation
n is the number of observations
∑ represents the summation symbol

3.4.3 *Root mean square error (RMSE)*

The average difference between a statistical model's predicted outcomes and the actual values is determined by the root mean square error (RMSE). The formula for RMSE is:

$$\text{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widetilde{Y}_i)^2} \quad (3)$$

Here,
RMSE is Root Mean Squared Error
Y_i is the actual value of the target variable for the i^{th} observation
$\widetilde{Y}_i$ is the predicted value of the target variable for the i^{th} observation
n is the number of observations
∑ represents the summation symbol

Python libraries like TensorFlow and StatsModels are used in Google Colab to create and build our LSTM and ARIMA models. The accuracy data may be gathered and examined once the models have been trained and validated. To take into account changes in market dynamics, it is crucial to compare the performance of LSTM and ARIMA models over a number of stocks or time periods. The efficacy of each model may then be evaluated using the accuracy data to discover which one performs better at stock market prediction. This is performed in our next section.

4. Results and discussions

The goal of this study is to predict closing stock market values. To train our data and determine the closing price, we used two ML models. These two models' results are reported in this section along with a brief discussion of them. First, each algorithm's specific results are discussed. Their training and test diagrams are initially shown, followed by presenting the RMSE, and MAE testing scores for each of them. After that, a comparative analysis of the performance scores of both algorithms is performed. Although LSTM was the first algorithm we tried, ARIMA yields the best outcomes. Finally, a comparison with previous relevant work is also presented which shows the novelty of our work.

4.1 *LSTM model results*

After applying the LSTM technique to train our data, we obtain the predicted diagram, RMSE, and MAE values that represent the performance of the suggested model.

In Figure 2 we have shown the predicted graph of our train dataset 'ApexfoodsReverse'. This shows the predicted diagram from 2011 to 2021 (till November). The proper flow of a stock market price is clearly depicted in the training diagram, with discernible patterns of rise and fall accurately represented. This visualization effectively captures the dynamic nature of stock prices, showcasing fluctuations over time as the market responds to various factors and influences. The rise and fall of stock prices are evidently reflected in the diagram, providing valuable insights into the underlying trends and behaviors of the market.

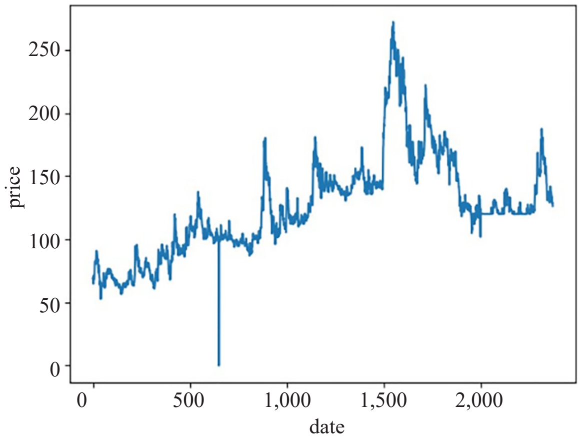

Figure 2. Apex Foods Train Diagram for LSTM Model

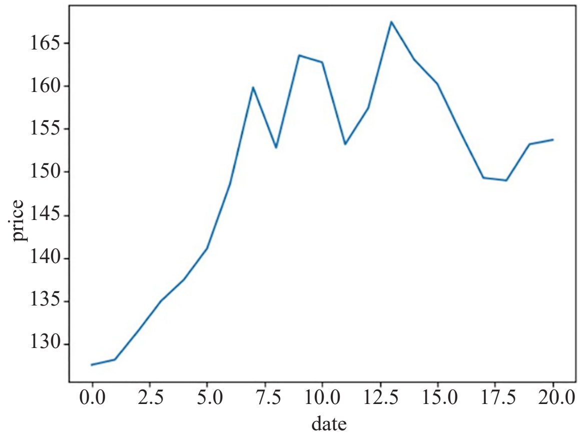

Figure 3. Apex Foods December 2021 Test Diagram for LSTM Model

In Figure 3, we have shown the predicted graph of our test dataset 'ApexfoodsDecember2021'. This shows the predicted diagram of Apex Foods for only December 2021. When compared to our train dataset diagram (Figure 2), we can see that it follows a similar pattern. The similarity between our train dataset diagram and the test dataset pattern indicates that our model has effectively learned from the training data. This alignment suggests that our model has captured the underlying structure and dynamics of the stock market prices present in the training dataset. Consequently, this learning capability enables our model to make reasonable predictions for new, unseen, test data. The consistency between the training and testing dataset diagram instills confidence in the model's ability to generalize its learnings and accurately forecast stock market prices for future instances.

In Figure 4 we have shown the predicted graph for both our train and test datasets. The figure shows the predicted values from 2011 to 2021 (till December). In this diagram, the orange-colored highlighted curves represent the outcomes of the training data, while the green-colored indicated curves represent the outcomes of the test data. As seen from the figure, the training and test diagrams align perfectly well with each other. This alignment indicates that the model performs effectively not only on the training data but also on new, unseen test data. Such consistency suggests that the model has successfully learned the underlying patterns in the training data and can generalize its predictions to new instances accurately. This robust performance across both training and test datasets enhances the reliability and confidence in the model's predictive capabilities, making it a valuable tool for stock market prediction tasks.

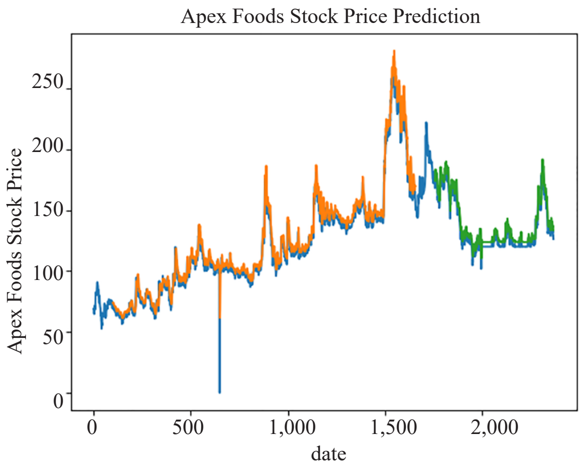

Figure 4. Apex Foods predicted Diagram from 2011-2021 for LSTM Model

In Table 2, we have presented the MAE and RMSE scores of the LSTM model. The LSTM model was trained on 70% of the data and tested on the remaining 30%. The test results reveal a Test RMSE of 141.45 and a Test MAE of 139.96. These metrics indicate the level of deviation between the predicted and actual stock prices, with lower values indicating better predictive accuracy. Overall, the LSTM model demonstrates reasonable predictive performance, with the RMSE and MAE values providing insights into the model's ability to generalize well to unseen data and accurately forecast stock market trends.

Table 2. MAE and RMSE Scores of LSTM Model

Algorithm	Training Data	Testing Data	Test RMSE	Test MAE
LSTM	70%	30%	141.45199017262152	139.95950777559275

4.2 *ARIMA model results*

After applying the ARIMA technique to train our data, we obtain the predicted diagram, MAE, and RMSE values that represent the performance of the suggested model.

The same as in LSTM, for ARIMA, we initially displayed the predicted graph of our train dataset, "Apex foods Reverse," in Figure 5. The forecast diagram from 2011 to 2021 (through November) is displayed below. The training diagram accurately depicts the rise and fall of stock market prices, showing clear patterns of movement. This visual representation effectively shows the dynamically shifting nature of stock prices, demonstrating variations over time in reaction to various market events. The diagram clearly depicts the rise and fall of stock values and provides insightful information about the underlying patterns and actions of the market.

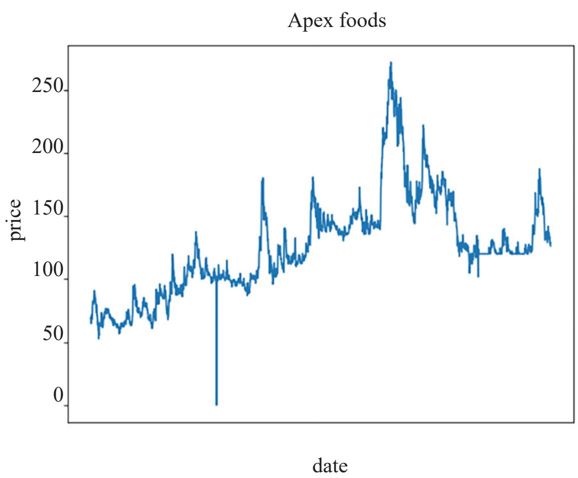

Figure 5. Apex Foods Train Diagram for ARIMA Model

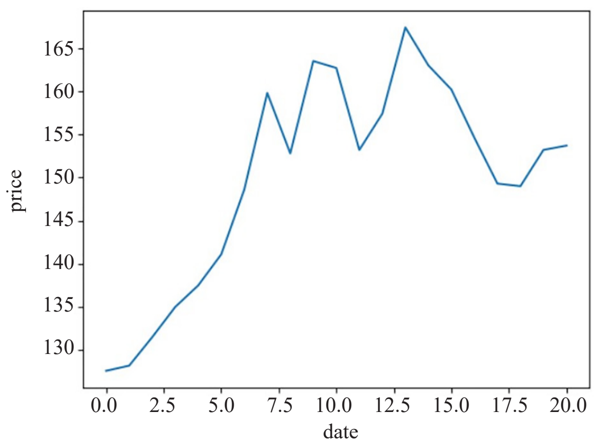

Figure 6. Apex Foods December 2021 Test Diagram for ARIMA Model

The anticipated graph for our test dataset, "Apex foods December 2021," is shown in Figure 6. This depicts the anticipated Apex Foods diagram for only December 2021. The similarity between the train and test dataset diagrams highlights how well the model learns from the training set. This correlation suggests that our model has been able to effectively understand the underlying dynamics and structure of the training dataset's stock market prices. Consequently, it has the ability to generate plausible forecasts for new, unseen test data. This graphical consistency between the training and testing datasets gives rise to trust in the model's ability to extrapolate its learned information and produce reliable stock market price estimates for future instances.

In Figure 7, we have shown the final predicted graph of our train dataset 'Apex foods Reverse'. The results of the training data are represented by the blue-highlighted curves in Figure 7, while the results of the test data are represented by the green-highlighted curves. The training and test diagrams exhibit identical patterns, indicating the model's effectiveness in performing well on both the training and unseen test data. This uniformity implies that the model has adeptly captured the underlying patterns within the training data and can accurately generalize its predictions to new instances. This consistency shows how well the model performs on both datasets, which increases trust in its ability to forecast outcomes for stock market tasks. This reliability enhances the model's value as a dependable tool for stock market prediction attempts.

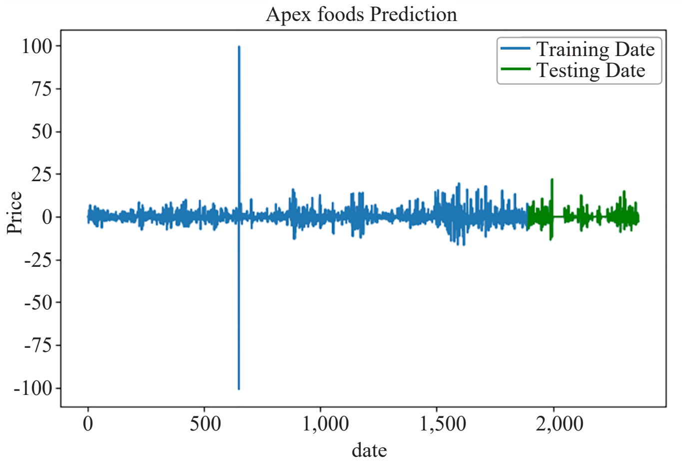

Figure 7. Final visualization of train and test data for ARIMA Model

Table 3 illustrates the performance metrics of our ARIMA model utilized for stock market prediction, with the model trained on 70% of the data and evaluated on the remaining 30%. The test results showcase a Test RMSE of 4.336 and a Test MAE of 3.45926. These metrics indicate the level of disparity between the predicted and actual stock prices, with lower values suggesting higher predictive accuracy. The ARIMA model demonstrates notably superior performance compared to the LSTM model, as evidenced by substantially lower RMSE and MAE values. This implies that the ARIMA model exhibits better predictive capabilities and precision in forecasting stock market trends.

Table 3. MAE and RMSE Scores of ARIMA Model

Algorithm	Training Data	Testing Data	Test RMSE	Test MAE
ARIMA	70%	30%	4.336	3.45926

4.3 *Comparison of results*

We have made use of RMSE and MAE to evaluate the performance of our two different approaches. Applying the LSTM method, we obtain test RMSE and MAE of 141.45199017262152 and 139.95950777559275. Similarly applying the same train and test data, we yield test RMSE and MAE values of 4.336 and 3.45926 by implementing the ARIMA method, respectively. These comparative values are presented in Table 4.

According to the aforementioned findings, the ARIMA has minimal values for RMSE and MAE. As a result, ARIMA offers a more accurate prediction than LSTM. Therefore, consumers may consequently readily comprehend the stock market's activity and form forecasts utilizing the approach suggested in this paper.

Table 4. RMSE and MAE Comparison of LSTM and ARIMA

Algorithm	Training Data	Testing Data	Test RMSE	Test MAE
LSTM	70%	30%	141.45199017262152	139.95950777559275
ARIMA	70%	30%	4.336	3.45926

4.4 *Comparison of our work with previous similar works*

Using the LSTM and ARIMA models, we computed the RMSE and MAE values using the Apex Foods data. Now that the RMSE and MAE values have been generated using the LSTM and ARIMA models, we will compare our work with some other similar works. This will highlight the novelty of our work. Here, we have considered those works where RMSE and MAE are used for performance parameters.

Table 5 provides a comparison of our research findings with those of previous works in the field of stock market prediction. Previous studies (21), (34), and (35) utilized LSTM and ARIMA algorithms, respectively. The performance scores reported in these studies include an RMSE of 390.46 for LSTM in (21), an RMSE of 5.999 for ARIMA in (34), and an MAE of 23.89 for ARIMA in (35). In contrast, our study achieved improved performance metrics, with an RMSE of 141.45199 an MAE of 139.9595 for LSTM, an RMSE of 4.336, and an MAE of 3.45926 for ARIMA. These results demonstrate the superior predictive accuracy of our LSTM and ARIMA models compared to those reported in previous works.

Table 5. Comparison of Our Work with Previous Works

Research Work	Utilized Algorithm	Performance Score
(21)	LSTM	RMSE = 390.46
(34)	ARIMA	RMSE = 5.999
(35)	ARIMA	MAE = 23.89
This Work	LSTM	RMSE = 141.45199 MAE = 139.9595
	ARIMA	RMSE = 4.336 MAE = 3.45926

5. Conclusion and future work

In conclusion, this paper contributes to the field of stock market forecasting by offering a comprehensive analysis of LSTM and ARIMA models, providing valuable insights into their predictive capabilities, and setting a new benchmark for stock price prediction. The findings advocate for a balanced perspective in model selection, emphasizing the importance of considering traditional statistical models alongside advanced machine learning techniques. As the quest for more accurate and reliable stock market predictions continues, this research serves as a testament to the evolving landscape of financial analytics, urging future studies to build upon these findings and further refine predictive methodologies for the financial domain.

The study's findings, while substantial, open the door to several opportunities for refinement, innovation, and application. Future studies can explore the integration of LSTM and ARIMA models, harnessing the strengths of both deep learning and traditional statistical methods. Another future scope might be expanding the dataset to include alternative predictors such as economic indicators, news sentiment analysis, or social media trends might enhance the model's predictive capabilities. In addition, implementing the models in a real-time prediction framework could be a significant step forward. Finally, exploring more sophisticated deep learning architectures such as Transformer models or Gated Recurrent Units (GRUs) could uncover new insights. By exploring these avenues, future research based on this paper can significantly contribute to the field of stock market predictions.

Conflict of interest

The authors declare no competing financial interest.

Author contributions

Ishtiaq Ahammad: Supervision, conceptualization, investigation, data curation, Draft manuscript preparation, writing-review and editing. William Ankan Sarkar, Famme Akter Meem, Jannatul Ferdus: Data collection, conceptualization, model design, model training, analysis and interpretation of results, investigation on challenges, writing-original draft, Draft manuscript preparation. Md. Kawsar Ahmed, Md. R. Rahman, Rabeya Sultana, Md Shihabul Islam: Data collection, analysis and interpretation of results.

References

[1] Raza S, Baiqing S, Khine PK, Kemal MA. Uncovering the effect of news signals on daily stock market performance: An econometric analysis. *International Journal of Financial Studies*. 2023; 11(3): 99.

[2] Tang Y, Song Z, Zhu Y, Yuan H, Hou M, Ji J, et al. A survey on machine learning models for financial time series forecasting. *Neurocomputing*. 2022; 512: 363-380. Available from: doi:10.1016/j.neucom.2022.09.003.

[3] Rapach D, Zhou G. Forecasting stock returns. *Handbook of Economic Forecasting*. 2013; 2(Part A): 328-383.

[4] Hao PY, Kung CF, Chang CY, Ou JB. Predicting stock price trends based on financial news articles and using a novel twin support vector machine with fuzzy hyperplane. *Applied Soft Computing*. 2021; 98: 106806. Available from: doi:10.1016/j.asoc.2020.106806.

[5] Aldoseri A, Al-Khalifa KN, Hamouda AM. Re-Thinking data strategy and integration for artificial intelligence: Concepts, opportunities, and challenges. *Applied Sciences*. 2023; 13(12): 7082.

[6] Hemal SH, Khan AR, Ahammad I, Rahman M, Khan AS, Ejaz S. Predicting the impact of internet usage on students' academic performance using machine learning techniques in Bangladesh perspective. *Social Network Analysis and Mining*. 2024; 14(1): 66.

[7] Hossain A, Ahammad I, Ahmed K, Ahmed MI. Prediction of the computer science department's educational performance through machine learning model by analyzing students' academic statements. *Artificial Intelligence Evolution*. 2023; 4(1): 70-87.

[8] Prosad R, Khan AR, Ahammad I. Design of class routine and exam hall invigilation system based on genetic algorithm and greedy approach. *Asian Journal of Research in Computer Science*. 2022; 13(3): 28-44.

[9] Rahman K, Khan AR, Ahammad I. Online support for education, medication, agriculture and relief work at COVID-19 pandemic time. *Asian Journal of Research in Computer Science*. 2022; 13(4): 65-85.

[10] Khan MAR, Akter J, Ahammad I, Ejaz S, Khan TJ. Dengue outbreaks prediction in Bangladesh perspective using distinct multilayer perceptron NN and decision tree. *Health Information Science and Systems*. 2022; 10(1): 1-18. Available from: doi:10.1007/s13755-022-00202-x.

[11] Khan AR, Afrin F, Prity FS, Ahammad I, Fatema S, Prosad R, et al. An effective approach for early liver disease prediction and sensitivity analysis. *Iran Journal of Computer Science*. 2023; 6(4): 277-295. Available from: doi:10.1007/s42044-023-00138-9.

[12] Robel, Khan AR, Ahammad I, Alam M, Hasan K. Cricket players selection for national team and franchise league using machine learning algorithms. *Cloud Computing and Data Science*. 2023; 5(1): 108-139.

[13] Klocker F, Bernsteiner R, Ploder C, Nocker M. A machine learning approach for automated cost estimation of plastic injection molding parts. *Cloud Computing and Data Science*. 2023; 4(2): 87-111.
[14] Khan AR, Uzzaman F, Ahammad I, Prosad R, Khan TZ, Ejaz S, et al. Stock market prediction in bangladesh perspective using artificial neural network. *International Journal of Advanced Technology and Engineering Exploration*. 2022; 9(95): 1397-1427.
[15] Sonkavde G, Dharrao DS, Bongale AM, Deokate ST, Doreswamy D, Bhat SK. Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. *International Journal of Financial Studies*. 2023; 11(3): 94.
[16] Parmar I, Agarwal N, Saxena S, Arora R, Gupta S, Dhiman H, et al. Stock market prediction using machine learning. *2018 first international conference on secure cyber computing and communication (ICSCCC)*. Jalandhar, India: IEEE; 2018. p.574-576.
[17] Sonkavde G, Dharrao DS, Bongale AM, Deokate ST, Doreswamy D, Bhat SK. Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. *International Journal of Financial Studies*. 2023; 11(3): 94.
[18] Nikou M, Mansourfar G, Bagherzadeh J. Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management*. 2019; 26(4): 164-174.
[19] Namasudra S, Dhamodharavadhani S, Rathipriya R, Crespo RG, Moparthi NR. Enhanced neural network-based univariate time-series forecasting model for big data. *Big Data*. 2024; 12(2): 83-99.
[20] Sirimevan N, Mamalgaha IGUH, Jayasekara C, Mayuran YS, Jayawardena C. Stock market prediction using machine learning techniques. *2019 International Conference on Advancements in Computing (ICAC)*. Thrissur, India: IEEE; 2019. p.192-197.
[21] Mehtab S, Sen J, Dutta A. Stock price prediction using machine learning and LSTM-based deep learning models. *Machine learning and metaheuristics algorithms, and applications*. Springer Singapore; 2021. p.88-106.
[22] Mehtab S, Sen J. A time series analysis-based stock price prediction using machine learning and deep learning models. *International Journal of Business Forecasting and Marketing Intelligence*. 2020; 6(4): 272-335.
[23] Sharma A, Bhuriya D, Singh U. Survey of stock market prediction using machine learning approach. *2017 International conference of electronics, communication and aerospace technology (ICECA)*. Thrissur, India: IEEE; 2017. p.506-509.
[24] Mankar T, Hotchandani T, Madhwani M, Chidrawar A, Lifna CS. Stock market prediction based on social sentiments using machine learning. *2018 international conference on smart city and emerging technology (ICSCET)*. Mumbai, India: IEEE; 2018. p.1-3.
[25] Khan W, Ghazanfar MA, Azam MA, Karami A, Alyoubi KH, Alfakeeh AS. Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*. 2022; 13(7): 3433-3456.
[26] Mintarya LN, Halim JNM, Angie C, Achmad S, Kurniawan A. Machine learning approaches in stock market prediction: A systematic literature review. *Procedia Computer Science*. 2023; 216(1): 96-102.
[27] Usmani M, Adil SH, Raza K, Ali SSA. Stock market prediction using machine learning techniques. *2016 3rd international conference on computer and information sciences (ICCOINS)*. Thrissur, India: IEEE; 2016. p.322-327.
[28] Subasi A, Amir F, Bagedo K, Shams A, Sarirete A. Stock market prediction using machine learning. *Procedia Computer Science*. 2021; 194: 173-179.
[29] Kompella S, Chilukuri KC. Stock market prediction using machine learning methods. *International Journal of Computer Engineering and Technology*. 2019; 10(3): 20-30.
[30] Vijh M, Chandola D, Tikkiwal VA, Kumar A. Stock closing price prediction using machine learning techniques. *Procedia Computer Science*. 2020; 167: 599-606. Available from: doi:10.1016/j.procs.2020.03.326.
[31] Kumar D, Sarangi PK, Verma R. A systematic review of stock market prediction using machine learning and statistical techniques. *Materials Taday: Proceedings*. 2022; 49(8): 3187-3191.
[32] Srivinay, Manujakshi BC, Kabadi MG, Naik N. A hybrid stock price prediction model based on PRE and deep neural network. *Data (Basel)*. 2022; 7(5): 51.
[33] Haque M. *Dhaka stock exchange: Historical prices of stocks.* Kaggle; 2023. Available from: https://www.kaggle.

com/datasets/mahmudulhaque/dsebd/data [Accessed 2ed May 2024].

[34] Siami-Namini S, Tavakoli N, Namin AS. A comparison of ARIMA and LSTM in forecasting time series. *2018 17th IEEE international conference on machine learning and applications (ICMLA)*. Orlando, FL, USA: IEEE; 2018. p.1394-1401.

[35] Zhou K, Wang WY, Hu T, Wu CH. Comparison of time series forecasting based on statistical ARIMA model and LSTM with attention mechanism. *Journal of Physics: Conference Series*. 2020; 1631(1): 012141.