Stock Market Price Prediction using LSTM

Ramandeep Singh
Dept. of Computational Intelligence,SRM College of Science and Technology® Kattankulathur-603203

Sarthak Sharma
Dept. of Computational Intelligence,SRM College of Science and Technology® Kattankulathur-603203

Dr.A.Saranya
Dept. of Computational Intelligence,SRM College of Science and Technology® Kattankulathur-603203

***Abstract*—The term "stock market" often refers to a group of exchanges and locations where buyers and sellers come together to trade equity shares of publicly traded companies. The stock market is a part of a free-market economy since it gives alltypes of investors access to trading and capital exchange. Your entire investment might be lost in the stock market. Investorswill sell a company's shares if it performs poorly, which will cause the stock price to fall. You will not recover your initial investment when you sell. To address this, a range of Neural Network prediction techniques and algorithms were compared and examined. The pipeline works with the LSTM (Long Short- Term Memory) algorithm are more effective and aid in providingmore prediction accuracy. This project's objective is to forecast the stock price at various points in time, taking into account boththe accuracy and efficiency of the market price.**

I. INTRODUCTION

One of the crucial tasks in the financial sector is trading. One component of this industry is the stock market. It includes the future valuation of any stock and the trading of financial instruments on a financial exchange. When we forecast the stock market, it implies we're forecasting the movement of all the exchange's stocks in the future. To achieve this, historical analysis aids in forecasting future analysis of future stock values. Analysing the historical pattern of the equities to anticipate the market is highly complex and challenging. Due to its extreme volatility and daily fluctuations that occur at market opening. Customers of stock exchanges are constantly in need of technologies that can accurately forecast future stock prices. There is a big need for the software. For stocks, forecasting may be done to assess historical data as well as some future data. But it's hard to determine whether it's correct. Numerous topics are included in the forecasts, including as commerce and industry, environmental science, economics, and the financial sector. Short-term forecasting (in this category, the prediction can be done for seconds, minutes, days, weeks, or months), medium-term forecasting (this prediction is valid for 1 year or 2 years), and long-term forecasting (it is valid for more than 2 years) are the three categories into which forecasting can be divided. Time series data are analysed for the majority of forecasting problems. Data in a time series is a collection of observations of a chosen variable organised chronologically. We can determine the trends in the data, the pattern of the sequence, and the periods or cycle of the sequence to examine this time series data. If we look at the stock market, it is impossible to forecast whether it will be bullish or bearish, but with advance information, one may invest money intelligently in stocks and trade for market gains.

We may create a variety of patterns that illustrate the success of the firm through time with the use of various algorithms. Time series and forecasting are excellent study topics due of pattern analysis. This study discusses deep learning for stock price forecasting using the LSTM algorithm [5]. Deep learning is a technique that reads and trains through the given information, using the dataset's knowledge to predict the outcome accurately. In this study, deep learning with the LSTM algorithm is used to learn how to estimate stock values for big and small businesses as well as how to extract forecasts for short- and long-term frequencies from time series.

II. LITERATURE SURVEY

In a review of the literature, many types of algorithms were evaluated with the goal of forecasting or predicting the outcome for actual stock value [6]. The article proposes a strategy for various stock price situations and aims to obtain results in real time. Stock price forecasting by applying reinforcement learning algorithm. Since Markov processes are what characterise stock price forecasting, the issue is approached as a Markov process, and it is optimised using a reinforcement learning algorithm that picks up forecasting skills via experience with various datasets. Artificial neural networks are used in this approach to determine the stock price trend within the specified time.

Stock price prediction and stock transaction prediction modelling and analysis based on long short term memory have always been difficult tasks for academics in the financial sphere. The stock's history data may be learned by ANN since the stock price, volume, and other data include a wealth of information that influences stock price fluctuations. However, in reality, there are numerous factors that impact financial data, which leads to more random time series that often have multi-level and multi-scale features. As a result, the single

neural network prediction model has drawbacks and can affect how accurately stock values are predicted. Prediction of stock price using LSTM and SVR. The technique of support vector regression is employed in this research study to forecast stock prices for the stick market. This method is used by researchers for very accurate financial market predictions. The LSTM approach effectively predicts stock price by saving the dataset for a long period as well as the data from the previous day. Input gate, forget gate, and output gate are the three gates that fall under the category of using LSTM to validate prior states. The SVR methodology, which forecasts the outcome of forecasting using time series data, is another effective method for working with time series data. Here we cover many meth- ods of stock market forecasting and prediction using machine learning and other techniques for comparative comparison. Additionally, it demonstrates the PSO-LSTM neural network's forecasting or prediction of rising stock prices. Following the introduction of artificial intelligence technology, the stock market is added to a new generation of study due to forecasting and precise prediction. These days, a technology that is widely used to anticipate the financial time series sequence is machine learning. It keeps track of the highest price and forecasts the stock's eventual price. Particle optimisation model can be added to the current model to enhance the prediction outcome. With the use of empirical research, the LSTM model and PSO optimisation method may quickly and accurately determine the necessary network weight for the neural network, as well as lower the loss function and accomplish speedy fitting [10]. Stock price can be predicted using LSTM neural networks, and this theory was developed in response to the need for stock price forecasting.

III. METHODOLOGY

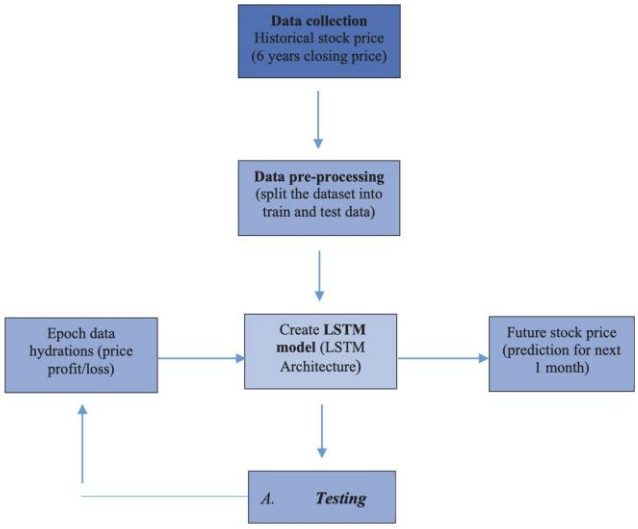


Fig. 1. Above figure shows the methodology of Stock price prediction using LSTM

A. Data Collection

Datasets are gathered through the pandas-data reader website , using the documentation for the programme. Additionally, there are several libraries where API keys may be used to gain access to a variety of datasets. The API key is made available in pandas-data reader so that users may easily retrieve the necessary stock dataset from the web server. Using the API key, you may make fifty requests each day to obtain the desired stock data. The open price, closing price, high, low, and dates from the recent 2017 to the present (six years) are included in this project's dataset. In the study, we used closing price information to forecast future prices.

Date	Open	High	Low	Close	Adj Close	Volume	company_name
2023-01-17 00:00:00-05:00	98.680000	98.889999	95.730003	96.050003	96.050003	72755000	AMAZON
2023-01-18 00:00:00-05:00	97.250000	99.320000	95.379997	95.459999	95.459999	79570400	AMAZON
2023-01-19 00:00:00-05:00	94.739998	95.440002	92.860001	93.680000	93.680000	69002700	AMAZON
2023-01-20 00:00:00-05:00	93.860001	97.349998	93.199997	97.250000	97.250000	67307100	AMAZON
2023-01-23 00:00:00-05:00	97.559998	97.779999	95.860001	97.519997	97.519997	76501100	AMAZON
2023-01-24 00:00:00-05:00	96.930000	98.089996	96.000000	96.320000	96.320000	66929500	AMAZON
2023-01-25 00:00:00-05:00	92.559998	97.239998	91.519997	97.180000	97.180000	94261600	AMAZON
2023-01-26 00:00:00-05:00	98.239998	99.489998	96.919998	99.220001	99.220001	68523600	AMAZON
2023-01-27 00:00:00-05:00	99.529999	103.489998	99.529999	102.239998	102.239998	87678100	AMAZON
2023-01-30 00:00:00-05:00	101.089996	101.739998	99.010002	100.550003	100.550003	70566100	AMAZON

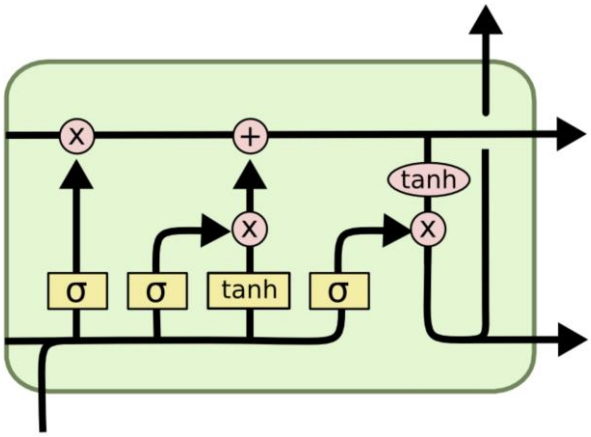
B. Data pre-processing

There are several methods for training and testing data during the pre-processing stage of data. Try to split the dataset into two parts, such as train data and test data, when dealing with time series data. Prediction in time series and sequence of datasets is always based on the pricing of the prior datasets. In this instance, the training dataset should make up 65% of the entire data frame length, and the size of the test dataset should be equal to the data frame length minus the training dataset. When processing data beforehand, taking into account the timesteps is helpful in computing the output for the following day, which relies on how many days went by before. It basically means to calculate the previous input to predict the future output, so it will start with the zeroth hydration and continue to run and analyse until it completes with the hundred hydrations. The process of data preprocessing is converting the given values into independent and dependent features based on the timesteps value. The hydration process aids in continuing to explore the specifics of a particular dataset value.

IV. LSTM MODEL

Accuracy is a key component of the stock market. The majority of researchers work really hard to acquire the highest accuracy; as a result, all approaches are accessible for this goal. To get the greatest results, it is our responsibility to choose the most accurate algorithms. To accomplish this goal, the optimal method must be used, such as logistic regression, long shot term memory (LSTM), support vector machine (SVM), random forest, etc. Because of its capacity for short- term prediction, LSTM is the best algorithm among those that are now accessible. The method that can recognise order dependency in sequence problems is called LSTM. The steps in the LSTM process include training the dataset, process- ing the data in accordance with past records, receiving the algorithm's result, and then comparing the result with many performance metrics. After accounting for all of these possi- bilities, the result is more precise and better. The approach is mostly used for the sequencing issue, and its prediction has been shown to be effective. The LSTM algorithm effectively maintains historical data while forgetting information that is

not currently available. It has a lot of historical data stored and is particularly effective at analysing time series datasets.



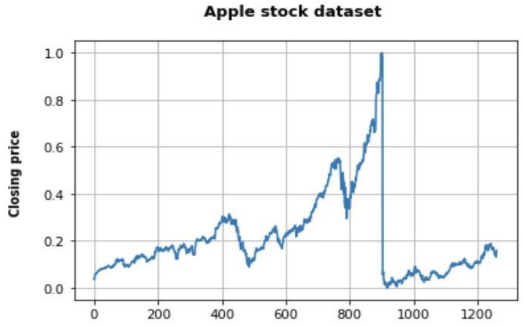
An internal cell state function serves as long short-term memory in LSTM. It is referred to as transient memory in cells. When it's necessary to rely on older inputs rather than the most recent, this trait might be helpful for reading, analysing, and making predictions. The LSTM employs a number of "gates" that regulate how data is entered into, stored in, and exited from the network. Using the sequential model to create the stacked LSTM, the first LSTM layer adds the value fifty and returns a true value, and the input shape is the value of the train and test data. Stack another layer of the LSTM model after that. Mean squared errors may be compiled with the use of LSTM.

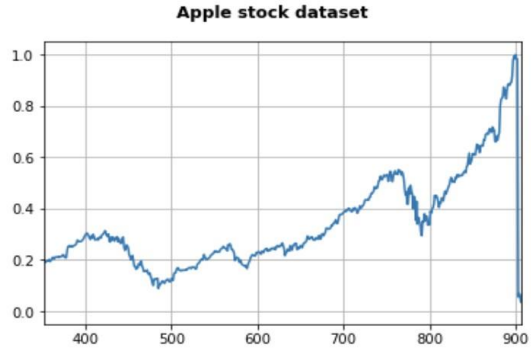
V. EXPERIMENTAL RESULTS

This experiment was conducted using a Python-based Jupyter Notebook. The libraries pandas, numpy, and matplotlib are utilised in this project. Dataset was obtained from the pandas-data reader website. Using the Tiingo libraries, the required dataset was selectively collected using the API key. Using the API key, a direct call to the web server was made to obtain the necessary stock dataset. The data pre-processing is done by utilising the time series approach, partitioning the dataset into different ways of manner, and in this instance, the dataset have contain the open price, closing price, high, low, dates from the past 2017 to current (six years). Divide the dataset into train and test data, then take into account the time steps value to compute the results for the following day depending on how many pre-processing steps were completed for the previous dataset. Consider the timesteps value in this instance to be 100. This basically means that there will be 100 features of data, and they will be converting this data into dependent and independent features based on the timesteps value given as 100. Timesteps will essentially indicate to calculate the values that are taken into consideration by using the input from the prior 100 days. Additionally, in order to generate the stacked LSTM, the sequential model and dense will be used. The LSTM algorithm was used to accomplish the stock dataset's data analysis. With the aid of pyplot and matplotlib, future stock price predictions are depicted visually through graphs.

VI. RESULT AND VISUALIZATION

The stock price dataset for the years between 2017 and 2022 (up to the present) is displayed in the following figures using the matplotlib software. It displays the closing price of the 'AAPL' stock using data from the supplied dataset with an index value of 1259 rows.

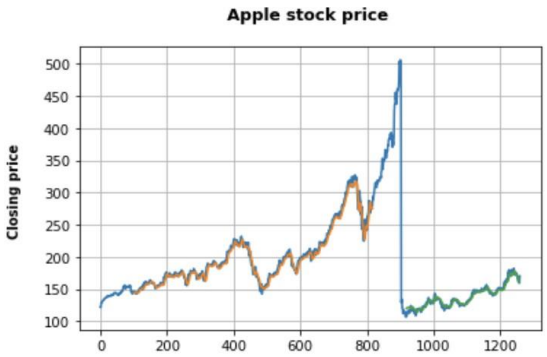


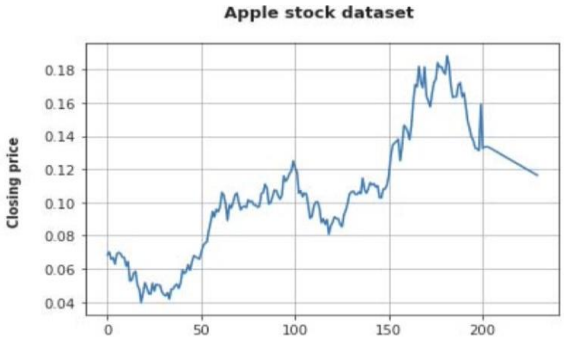


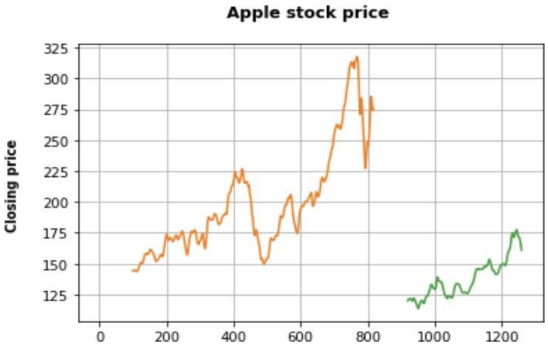
The process of inserting train and test data values into the LSTM model to produce the output and analyse the data by utilising the hydration technique in epochs is shown in this data analysis and hydration visualisation.

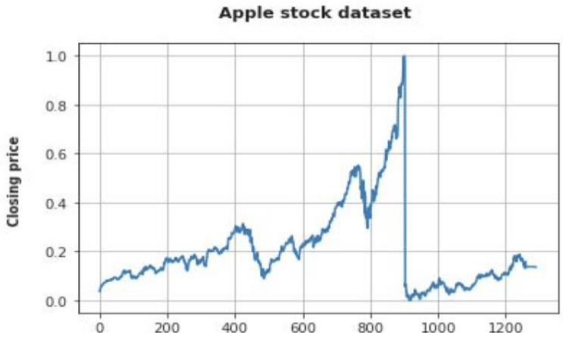
```
Epoch 1/15
45/45 [==============================] - 15s 124ms/step - loss: 0.0275
Epoch 2/15
45/45 [==============================] - 7s 152ms/step - loss: 0.0061
Epoch 3/15
45/45 [==============================] - 5s 121ms/step - loss: 0.0049
Epoch 4/15
45/45 [==============================] - 7s 151ms/step - loss: 0.0051
Epoch 5/15
45/45 [==============================] - 6s 130ms/step - loss: 0.0044
Epoch 6/15
45/45 [==============================] - 7s 145ms/step - loss: 0.0043
Epoch 7/15
45/45 [==============================] - 5s 118ms/step - loss: 0.0045
Epoch 8/15
45/45 [==============================] - 7s 147ms/step - loss: 0.0039
Epoch 9/15
45/45 [==============================] - 6s 123ms/step - loss: 0.0038
Epoch 10/15
45/45 [==============================] - 7s 151ms/step - loss: 0.0034
Epoch 11/15
45/45 [==============================] - 6s 121ms/step - loss: 0.0037
Epoch 12/15
45/45 [==============================] - 6s 129ms/step - loss: 0.0033
Epoch 13/15
45/45 [==============================] - 6s 142ms/step - loss: 0.0031
Epoch 14/15
45/45 [==============================] - 6s 129ms/step - loss: 0.0029
Epoch 15/15
45/45 [==============================] - 6s 139ms/step - loss: 0.0030
<keras.callbacks.History at 0x7f9ec6565c40>
```

Following the development of the LSTM model, the graph in Figure 1.3 below displays the accuracy and comparison of prediction values for the training and test sets of data in relation to the historical stock closing price dataset (Fig. 1.1). An orange line in this graph depicts the forecast for the training data, and a green line the forecast for the value of the test data.

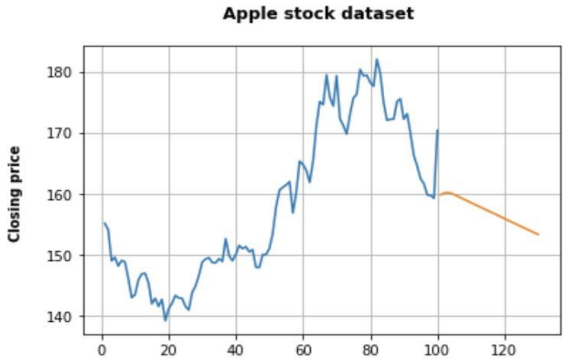


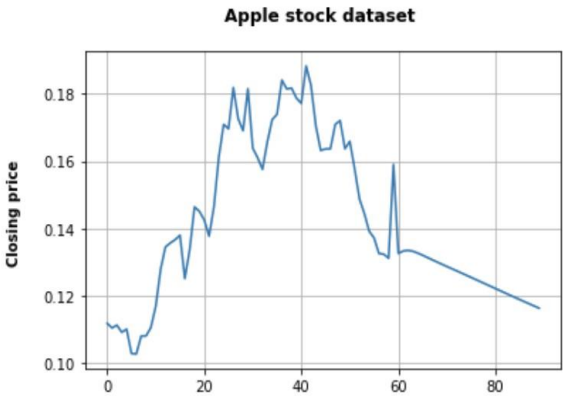






Last but not least, future prediction visualisation for the following 30 days using LSTM algorithms that train the model from the train and test data value along with scalar and inverse transform and carry out dataset analysis using epochs. The expanded orange line in figure 1.4, which depicts the final product, represents the price forecast for the coming month. The combined prediction value with the prior graph for the clear and ideal perspective is evident.





VII. CONCLUSION

The study for this article concentrated on stock market forecasting, which seeks to anticipate future changes in the stock value of a financial exchange. Investors will be able to earn more if share price movements can be predicted accurately. In order for the model to be approximated more quickly and with less error, the LSTM algorithm, a small batch gradient descent method, was employed to optimise the model in the manner of tiny step iteration used by small batch gradient descent algorithm. In contrast to previously used deep learning models, this experiment shows how the LSTM model can forecast stock value more accurately.

REFERENCES

1. Julee Kumari; Vivek Sharma; Sansar Chauhan ”Prediction of Stock Priceusing Machine Learning Techniques: A Survey”2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)
2. Pushpendra Singh Sisodia; Anish Gupta; Yogesh Kumar; Gaurav Ku- mar Ameta ”Stock Market Analysis and Prediction for Nifty50 using LSTM Deep Learning Approach” 2022 2nd International Conferenceon Innovative Practices in Technology and Management (ICIPTM)
3. Samuel Olusegun Ojo; Pius Adewale Owolawi; Maredi Mphahlele; Juliana Adeola Adisa ”Stock Market Behaviour Prediction using StackedLSTM Networks” 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC)
4. Bandi Jaswanth; Jitendra Kaushik ”Stacked LSTM a Deep Learning model to predict Stock market” 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)
5. Nitin Sharma; Biju R Mohan ”Stock Market Prediction Using HistoricalStock Prices ” 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC)
6. G.W.R.I. Wijesinghe; R.M.K.T. Rathnayaka ”Stock Market Price Fore- casting using ARIMA vs ANN; A Case study from CSE” 2020 2nd International Conference on Advancements in Computing (ICAC)