Chapter 1
Introduction

Abstract-

The stock market is incredibly volatile, because stock values fluctuate due to several factors, making stock prediction is a difficult and extremely difficult and complementary task. With rapidly increasing in technology in stock price prediction, investors nowadays require fast and reliable information to make efficient methods to predict stocks. Stock market price prediction is a difficult task that generally requires a lot of human-computer interaction.

The primary goal of this model is to determine how well our machine learning algorithm can predict and train our data, as well as how increasing in number of epochs and data size will enhance our model. When compared to other stock price prediction model, this will give more accurate results. The accuracy of this model is tested with various data sizes and epochs, and the results are tabulated. This article mainly focuses on developing a model for predicting future stock market values using Recurrent Neural Networks (RNN) and, in particular, the Long-Short Term Memory model (LSTM).

Introduction

Stock price prediction is a difficult task that is modelled using machine learning to determine stock prices. For stock market forecasting, there are a variety of methodologies and tools available. The stock market is thought to be extremely dynamic and complicated. An accurate prediction of future prices may result in a higher profit ratio for investors who invest in stocks. According to the estimates, investors will be able to select equities that will provide a higher profit.

Various machine learning techniques have been used in stock market prediction over the years, but with the increased amount of data and the assumption of more accurate predictions, deep learning models are now being used, which have proven to be superior to traditional machine learning methods in terms of prediction accuracy and speed. The Long-Short-Term Memory (LSTM) Recurrent Neural Network, one of the most prominent deep learning models for stock market prediction, will be discussed in this article. In this task, we will use python libraries to automatically download historical stock data and fit an LSTM model to the data to predict future stock values. We'll deploy the Long-Short-Term Memory (LSTM) Recurrent Neural Network, which is one of the most widely used deep learning models for stock market prediction. We will use python modules to retrieve historical stock data and fit the LSTM model with varying epochs and data sizes to make more accurate model.

Chapter 2

Literature Survey

The primary goal of our literature analysis was to assess various algorithms and models to see if stock price forecasts could be made based on actual stock prices. We proceeded to review existing plans, analyse major difficulties, and better ourselves because we were unable to detect a possible change in this stock price estimate. We identified LSTM after conducting a quick search for popular solutions to these problems. After selecting to use the LSTM neural network to produce stock forecasts, data from yahoo finance is collected for the stock over a 5- to 25-year timeframe.

Related Work

Hossain et al. [1] developed a hybrid model based on LSTM and Gated Recurrent Unit (GRU) that worked on S&P historical time series data and extracted the data using Yahoo's yfinance API. When compared to individual LSTM or GRU layer forecasts, the hybrid model had an MAE of roughly 0.023 and produced substantially more accurate predictions.

For forecasting the Bombay Stock Exchange (BSE) Sensex data, Shah et al. [2] evaluated LSTM and Deep Neural Network (DNN). Both models were suitable for daily forecasts and had an RMSE of around 1%, while the LSTM model was shown to be more suitable for weekly predictions.

Zhang et al. [3] developed a new stock price trend prediction system that predicts stock prices. For data, it employs the TA- Lib open source library, and for analysis and prediction, it employs Random Forest, Imbalanced learning, and Feature selection. It has a 67.5 percent accuracy and a 3.7 percent standard deviation, making it more ideal for predicting stock values over a longer period of time, such as 30 to 40 days.

Creighton et al. [4] used a hybrid model of exponential smoothing, ARIMA, and backpropagation neural network (BPNN) on the S&amp 500 and S&P 400 daily closing indexes to complement linear and nonlinear predictions. It is based on the KoNstanz Information MinEr (KNIME) analytics platform and works well for predicting weekly data with less noise and more linear growth, but not for daily predictions. The model has a 45.1 percent directional accuracy, an MAE of 16.68, an MSE of 434.121, and an RMSE of around 20.836.

On the NSE data, Selvin et al. [5] employed three algorithms: LSTM, RNN, and Convolutional Neural Network (CNN) - Sliding Window model. The sliding window was 100 minutes long, with 90 minutes for information and 10 minutes for prediction. The results of CNN were found to be superior to those of LSTM and RNN. RNN had an MAE of 5.13 percent, LSTM had an MAE of 5.31 percent, and CNN had an MAE of

4.98 percent, according to the model. ARIMA, a linear model with an MAE of 29.87 percent, was outperformed by deep learning.

For predictions, Sadia et al. [6] used Random Forest and SVM on a historical dataset from Kaggle. The data was trained on OHCLV, trade, and value parameters, yielding an SVM accuracy of 78.7% and a Random forest accuracy of 80.8 percent.

Deepak et al. [7] took a different technique, using SVM with RBF kernel on the BSE Sensex dataset to forecast stock market situations for the coming week, day, and minute. Depending on the share considered, the model had an accuracy of 80 to 85 percent. Tab. 1 outlines the numerous stock market prediction algorithms used on various datasets, as well as their accuracy and mean errors.

Persio et al. [8] employed LSTM, GRU, and Multilayer RNN to detect short term ups and downs in the Google stock price data for financial forecasting. It was also subjected to the Adam optimization technique. For a 5-day data projection, the model had a 72 percent accuracy rate.

A Comparison of Different Prediction Techniques

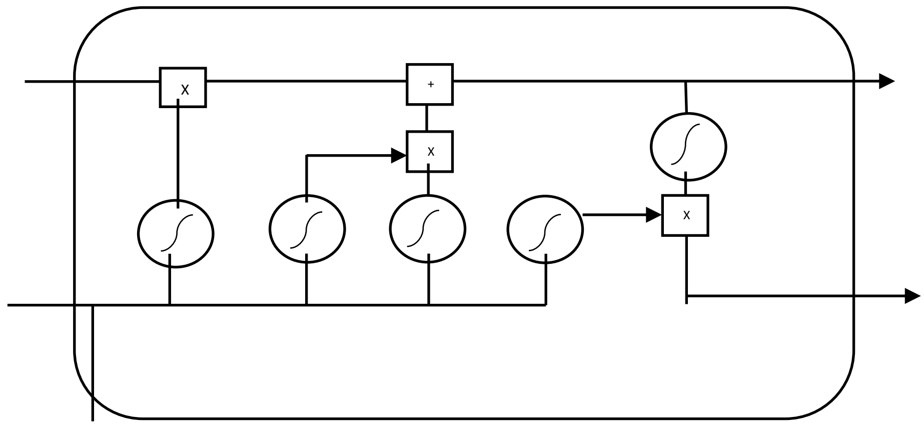
Paper Title	Author	Model Used	Evaluation
Hybrid Deep Learning Model for Stock Price Prediction	Hossain et al.	LSTM- GRU hybrid	MSE- 0.00098 MAE- 0.023
A Comparative Study of LSTM and DNN for Stock Market Forecasting	Shah et al.	LSTM, DNN	RMSE- 1%
A Novel Data-driven Stock Price Trend Prediction System	Zhang et al.	Random Forest	Accuracy- 67.5% Std deviation- 3.7%

Towards Building a Hybrid Model for Predicting Stock Indexes	Creighton et al.	ARIMA-BPNN hybrid	MAE- 16.68 MSE- 434.121 RMSE- 20.836 Accuracy- 45.1%
Stock price prediction using LSTM, RNN and CNN-sliding window model	Selvin et al.	RNN, LSTM, CNN	MAE - 5.13% (RNN) 5.31% (LSTM) 4.98% (CNN)
Stock Market Prediction Using Machine Learning Algorithms	Sadia et al.	SVM, Random Forest	Accuracy-78.7% (SVM) 80.8% (Random forest)
Machine Learning Approach in Stock Market Prediction	Raut Sushrut Deepak	RNN, LSTM	Accuracy- 80 to 85%
Recurrent Neural Networks Approach to the Financial Forecast of Google Assets	Persio et al.	GRU	Accuracy- 72%

Chapter 3
Proposed Model

We propose to use the LSTM (Long Short Term Memory) technique to give effective stock price prediction and to evaluate how well our model has been trained by considering various aspects such as epochs and data size.

LSTM – Overview



LSTMs are a type of RNN that may capture context-specific temporal dependencies over lengthy time periods. Each LSTM neuron is a memory cell that may store other data and keeps track of its own cell state. An LSTM neuron additionally takes in its old cell state and outputs its new cell state, whereas neurons in standard RNNs only take in their prior hidden state and the current input to output a new hidden state.

The following three components or gates, make up an LSTM memory cell, as shown in above figure:

1. Forget gate: This gate determines when particular elements of the cell state should be replaced with more recent data. It produces values that are near to 1 for sections of the cell state that should be kept and zero for values that should be ignored.

2. Input gate: This component of the network learns the conditions under which any information should be kept (or updated) in the cell state based on the input (i.e., previous output o(t-1), input x(t), and prior cell state c(t-1).

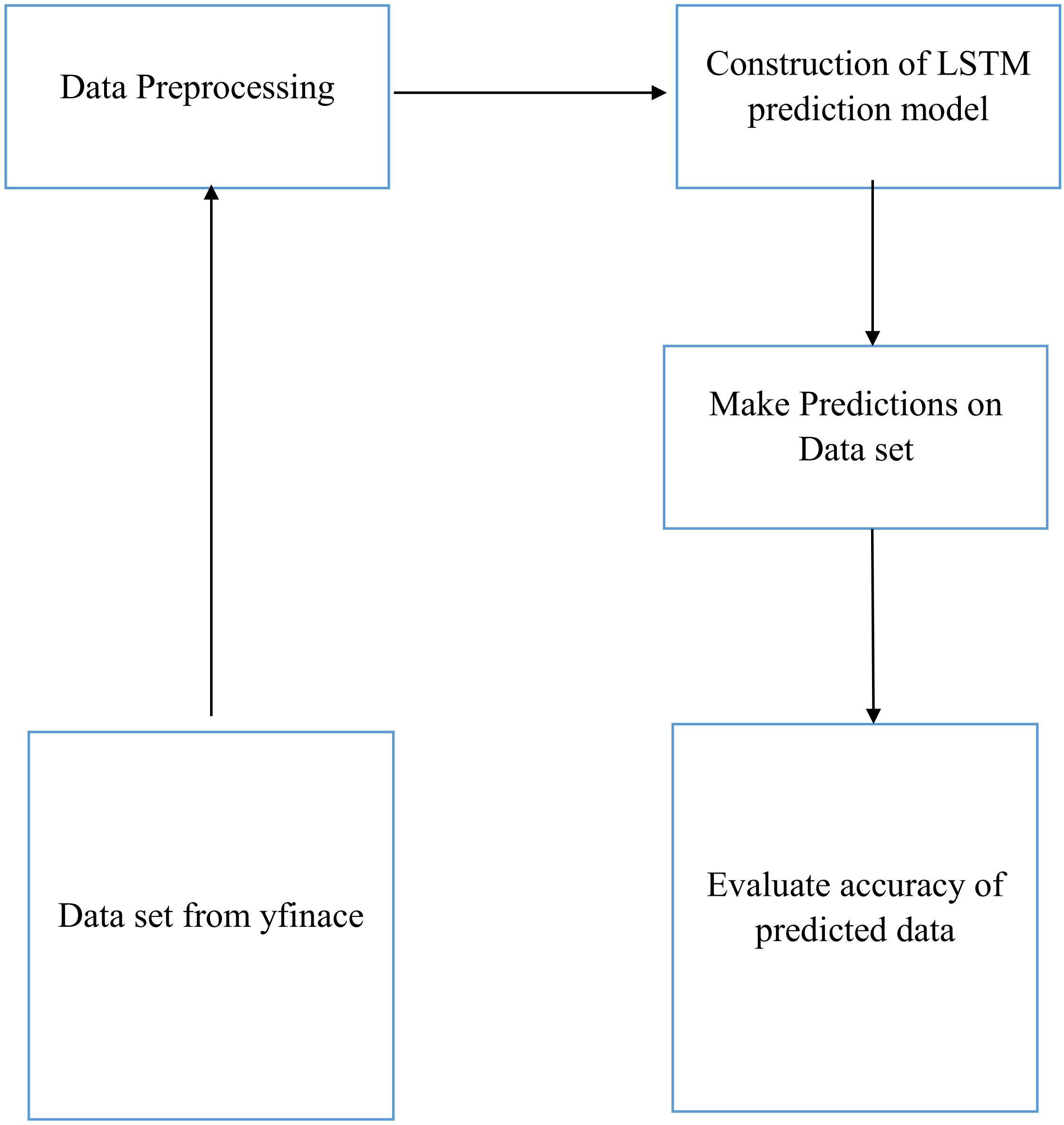
3. Output gate: This section determines what information is carried forward (i.e., output o(t) and cell state c(t)) to the next node in the network based on the input and cell state.

As a result, LSTM networks are suitable for examining how price changes in one stock affect the prices of numerous other stocks over time. They can also decide (dynamically) how long information about certain previous trends in stock price movement should be preserved in order to better predict future trends in stock price variation.

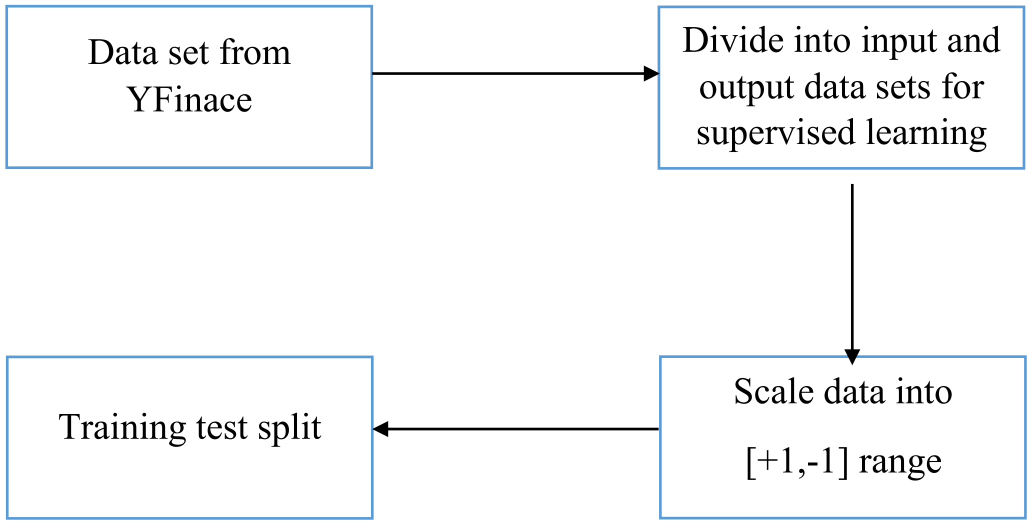
Advantage of LSTM

The ability of LSTM to read intermediate context is its key advantage. Without explicitly applying the activation function inside the recurring components, each unit remembers information for a long or short length of time. The release of the forget gate, which ranges between 0 and 1, is the only way for any cell state to be repeated. To put it another way, the LSTM cell's forgetting gateway is in charge of both the hardware and the function of cell state activation. As a result, instead of explicitly increasing or decreasing in each step or layer, the data from the preceding cell can flow through the unmodified cell, and the instruments can convert to their suitable values over a limited time. Because the amount stored in the memory cell is not transformed in a recurrent fashion, the gradient does not cease when trained to distribute rearward, allowing LSTM to solve a perishable gradient problem

System architecture



Data Pre-processing



Tiingo API, Yahoo, and Google Finance are all good places to look for stock market data. These websites provide APIs through which stock datasets from multiple firms can be downloaded by simply providing parameters.

The following processes are used to convert the data into a format that can be used with the prediction model:

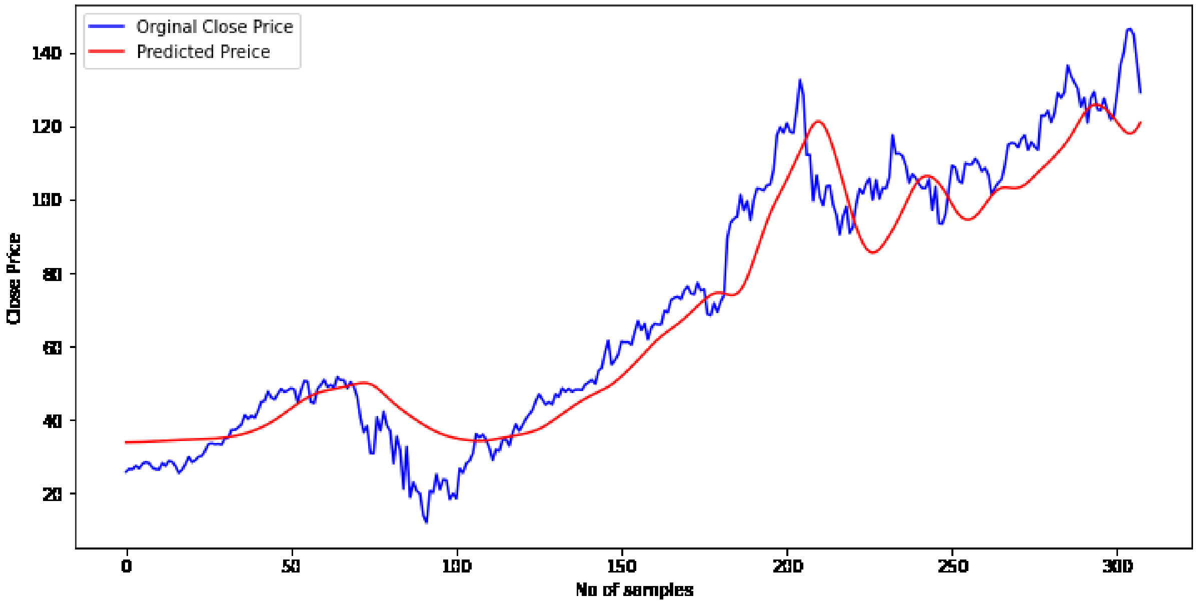
1. For supervised learning, transform time-series data into input-output components.

2. Changing the data's scale to [-1, +1].

Chapter 4
GUIs of Project

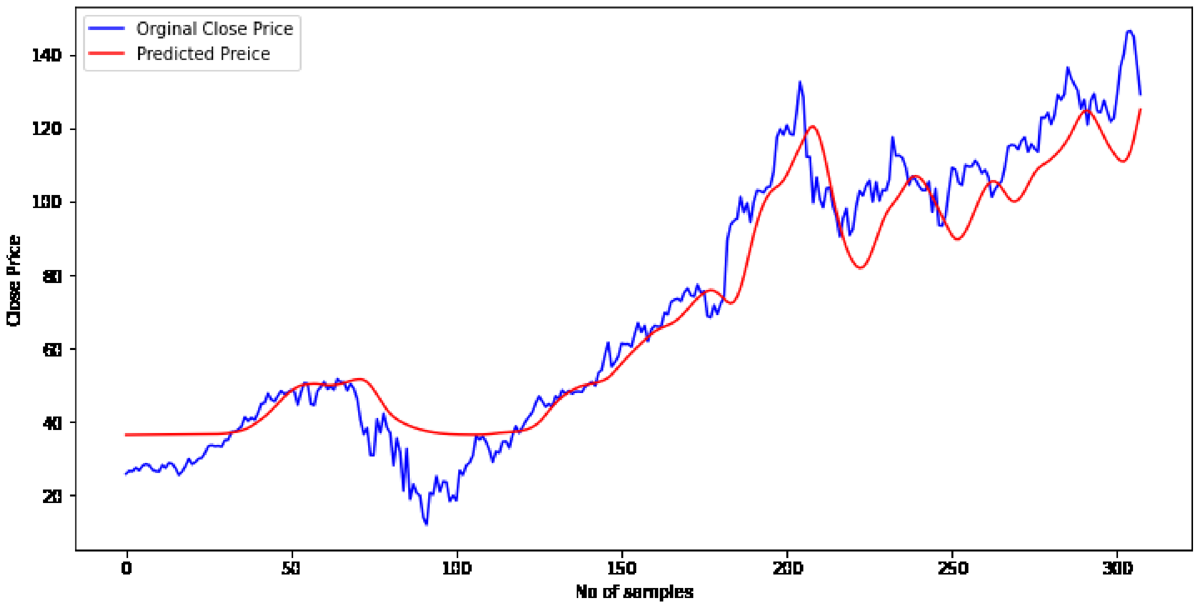
Comparison of Actual and Predicted Price with different Dataset size and epochs

Dataset size – 1025



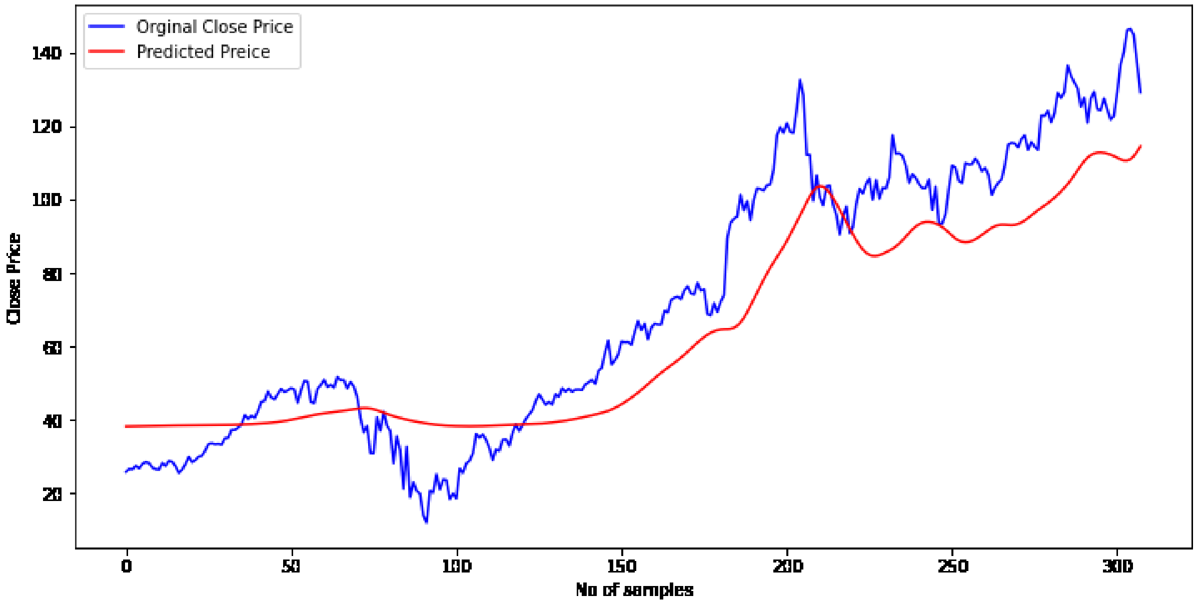
Dataset size - 1025

Epochs - 25



Dataset size - 1025

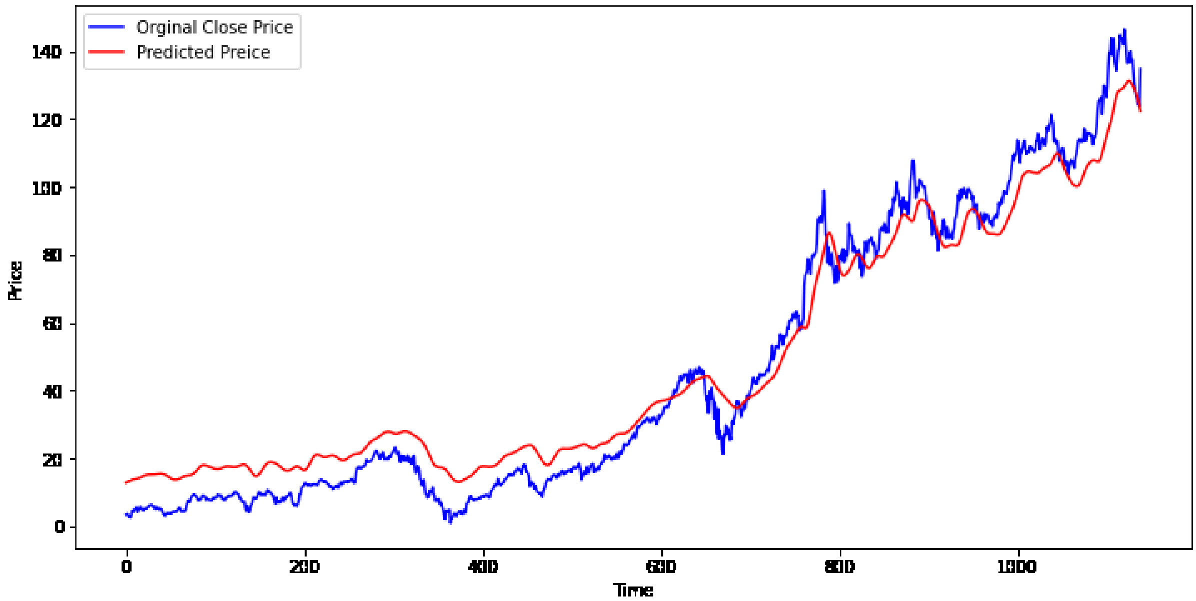
Epochs – 50



Dataset size - 1025

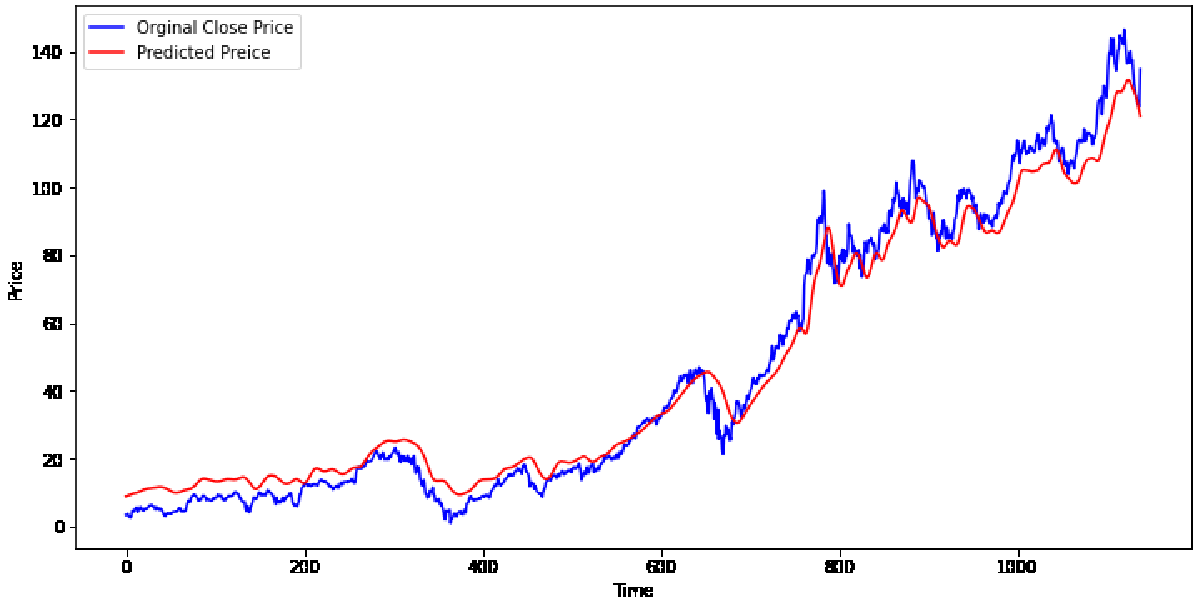
Epochs – 100

Dataset Size - 3795



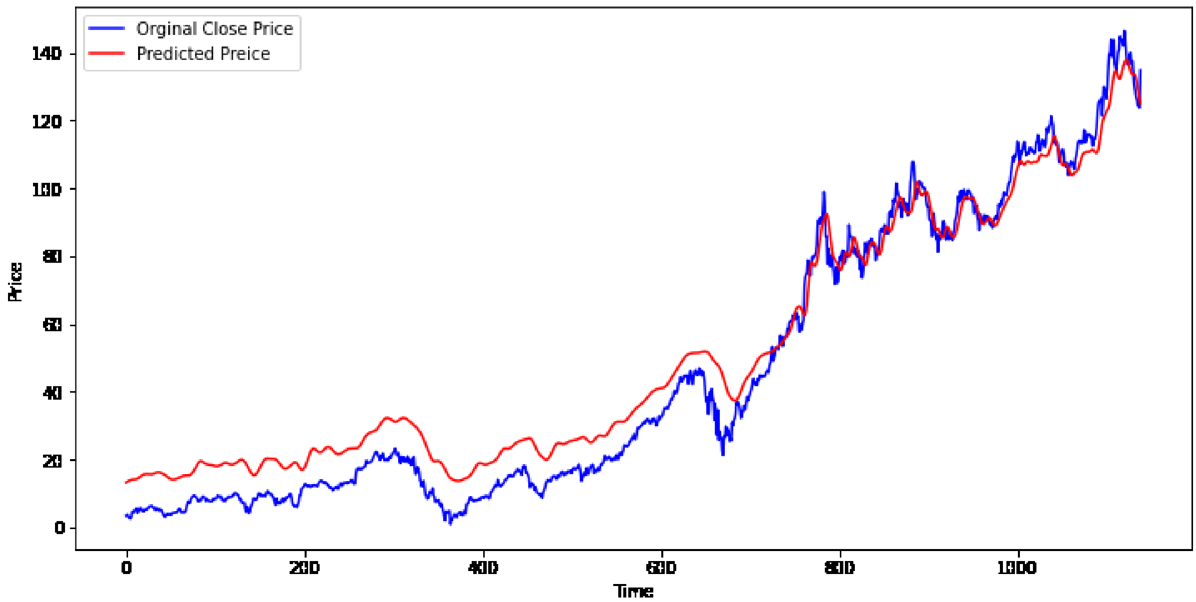
Dataset size - 3795

Epochs – 25



Dataset size - 3795

Epochs – 50



Dataset size - 3795

Epochs – 100

Result Analysis

Dataset size - 1025

	25 Epochs	50 Epochs	100 Epochs
Mean Absolute Error	8.5813	7.655	13.186
Mean Squared Error	112.680	102.796	245.170
Root Mean Square Error	10.615	10.138	15.657
R2 Score	0.916	0.923	0.818

Dataset size - 3795

	25 Epochs	50 Epochs	100 Epochs
Mean Absolute Error	7.261	5.213	7.42
Mean Squared Error	66.310	40.160	70.10
Root Mean Square Error	8.143	6.33	8.37
R2 Score	0.96	0.976	0.95

Conclusion

Many investors all over the world are interested in stock investing. Making a decision, on the other hand, is a difficult task due to the numerous factors involved. Investors are eager to forecast the stock market's future after making successful investments. Even a small improvement in performance can have a huge impact. By providing supporting information such as future stock price guidance, a good forecasting system can assist investors in making more accurate and profitable investments. Other factors, such as politics, economic growth, financial matters, and the atmosphere on social media, could influence prices in addition to historical prices. Emotional analysis has been shown in numerous studies to have a significant impact on future prices. As a result, combining technical and fundamental analysis can yield very accurate predictions.

This paper proposes the LSTM model for forecasting stock values. We can now say that if we take a large data set and train the LSTM model for 50 epochs, we will get highly accurate predicted data.

Our model's output has yielded some promising results. The results of the tests show that our model can track the evolution of closing prices.

We will try to find a method to forecast future data with the same accuracy in our future work, as well as develop a mobile application that shows all of the forecasting.

References

[1] Mohammad Asiful Hossain, Rezaul Karim, “Ruppa Thulasiram Hybrid Deep Learning Model for Stock Price Prediction”, IEEE Symposium Series on Computational Intelligence SSCI, pp.1837-1844 (2018).

[2] Dev Shah, Wesley Campbell, Farhana H. Zulkernine, “A Comparative Study of LSTM and DNN for Stock Market Forecasting”, IEEE International Conference on Big Data (2018)

[3] Jing Zhang, Shicheng Cui, Yan Xu, Qianmu Li, and Tao Li, “A Novel Data-driven Stock Price Trend Prediction System”, Elsevier Expert Systems with Applications, Vol. 97, pp.60–69 (2018).

[4] Jonathan Creighton, and Farhana H Zulkernine, “Towards Building a Hybrid Model for Predicting Stock Indexes ”, IEEE International Conference on Big Data (2017).

[5] Sreelekshmy Selvin, R Vinayakumar, E. A Gopalakrishnan, Vijay Krishna Menon, “Stock price prediction using LSTM, RNN and CNN-sliding window model”, IEEE Conference Paper, doi: 10.1109/ICACCI.2017.8126078, pp.1643- 1647 (2017).

[6] K. Hiba Sadia, Aditya Sharma, Adarrsh Paul, SarmisthaPadhi, Saurav Sanyal , “Stock Market Prediction Using Machine Learning Algorithms”, International Journal of Engineering and Advanced Technology, Vol. 8, No. 4, pp.25-31 (2019).

[7] Raut Sushrut Deepak, Shinde Isha Uday, and Dr D Malathi," Machine Learning Approach in Stock Market Prediction", International Journal of Pure and Applied Mathematics, Vol. 115, No. 8, pp.71-77 (2017).

[8] Luca Di Persio, and Oleksandr Honchar, "Recurrent Neural Networks Approach to the Financial Forecast of Google Assets", International Journal of Mathematics and Computers in Simulation, Vol. 11, pp.7-13 (2017).

[9] Usmani, Mehak, Syed Hasan Adil, Kamran Raza, and Syed Saad Azhar Ali, "Stock market prediction using machine learning techniques," 3rd international conference on computer and information sciences (ICCOINS), pp. 322-327. IEEE, 2016.

[10] Nayak, Aparna, MM Manohara Pai, and Radhika M. Pai, "Prediction models for Indian stock market" Procedia Computer Science 89 (2016): 441-449.

[11] Pahwa, Kunal, and Neha Agarwal, "Stock market analysis using supervised machine learning", 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). IEEE, 2019.

[12] Roondiwala, Murtaza, Harshal Patel, and Shraddha Varma, "Predicting stock prices using LSTM", International Journal of Science and Research (IJSR) 6.4 (2017): 1754-1756.