# Stock Market Prediction Using Deep Learning

Pooja[a], Nagalakshmi Vallabhaneni*[b], Harshini S R[c], Jayaprakash[d]
[a,c,d]Student
Master of Computer Applications, School of Information Technology and Engineering
Vellore Institute of Technology, Vellore - 632014, Tamil Nadu, India
[b]Assistant Professor

School of Information Technology and Engineering
Vellore Institute of Technology, Vellore - 632014, Tamil Nadu, India
E-mail : pooja.sharma2022@vitstudent.ac.in
nagalakshmi.v@vit.ac.in
harshini.sr2022@vitstudent.ac.in
jayaprakash.2022@vitstudent.ac.in

**ABSTRACT:**
A stock market is a place where people buy and sell their stocks and focuses on maximizing profit and minimizing their losses. And to achieve this, they try to predict the stock prices but, predicting the stock prices is a tedious process as changes are uncertain, but with some deep learning techniques, we can find hidden patterns in the stock value changes and predict their future values for better investments. Stock market values depend on more than one factor: the previous value of stocks, financial news related to companies, and social media trends, but most of the already existing work only considers one factor. This paper proposes a hybrid deep learning model (CNN-LSTM) and considers stock market data and sentiment analysis for prediction. We use the Convolutional Neural Network model (CNN) to classify the investors' sentiments about a stock as positive, negative, or neutral. We then predict the future values of stocks using the Long Short-Term Memory (LSTM) Neural Network model by combining direct stock market data with sentiment analysis data discussed above. At last, we measure the performance of our proposed model by comparing it with already existing baseline sentiment analysis models and stock market prediction models and evaluate its performance using statistical methods like RMSE, Accuracy, Precision, and recall.

**Keywords:** Stock Market Prediction, Hybrid model, Convolutional Neural Network, Long Short-Term Memory Neural Network, Sentiment Analysis.

**Introduction**:

The stock market has a reputation for being unpredictable, random, and volatile. It is a chaotic environment with an unbelievable amount of constantly changing data, which makes it challenging to anticipate the future and take profitable action based on those predictions. It truly belongs to the most difficult forecasting jobs for time series.The major objective of this capstone is to research and use deep learning techniques on the stock market to forecast stock behaviour and then act on those predictions to reduce investing risk and increase profit. Transfer learning will be used to accomplish the goal in order to benefit from neural network models that have already been created. Then, predictions are evaluated using real historical stock price data.In this research, we present a hybrid model for stock market analysis and price prediction by fusing a deep learning approach with a sentiment analysis model to get over these constraints. First, using a Convolutional Neural Network model to categorise the investors' hidden sentiments, we generate the sentiment factor. In addition, an LSTM neural network approach is used to predict the closing price one day in advance. Its inputs are investor sentiment from a prominent stock forum and technical indicators generated from stock quotes.

**Problem Description:**

In everyday life we come across the term stock market. If there is a reliable and trustworthy algorithm built for prediction of stocks then there will be higher opportunities with lower risk for people to invest more in the stock market. In this research, we propose a hybrid algorithm for stock market value prediction using deep learning concepts. Our model works with CNN algorithm for sentiment analysis using textual data and LSTM algorithm which then combines the results of CNN and technical stock data for prediction of stock values. For reliability and efficiency of our model we compare our model with already existing algorithms, in result it gives the least mean absolute percentage error for LSTM algorithm and prediction is almost close to the actual market price.

**Literature Survey**:

[1] In this research paper, extended model of RNN iDismisss applied such as LSTM,GRU.For stock price movement prediction and classification,they introduced a new BGRU model.Their proposed strategy may greatly outperform previous systems that merely employed historical price data in terms of stock prediction accuracy on a standard financial database, according to experimental results.

[2] Suggested utilizing social network analysis and economic news to predict advances in social media.Number of features from final data sets were implemented to boost the accuracy of forecasts.Prices in the capital markets have been implemented and are difficult to foresee.It is advised that social networking sites and financial forecasts have a greater impact on SMs.Implemented a comprehensive learning strategy for SM analysis.

[3] With statistical certainty, it is concluded that adding financial news feelings and stock attributes as input will considerably improve LSTM's performance in stock price forecasting and adding financial news feelings and stock attributes as input will considerably improve GRU's performance in stock price forecasting.

[4] The Long Short-Term Memory (LSTM) model and the Multilayer Perceptron (MLP) model are compared in this research. The process of feature engineering is one of the crucial processes in the construction of both models. The characteristics resulting from the heuristic analysis and study method used in this work are ideal for companies that are heavily traded. According to the findings, the MLP model has done better at predicting short-term stock values than the LSTM model. Neural networks have shown to be an effective tool for predicting a chaotic environment like the stock market.

[5] The ARIMA model is underperforming DL models. Since a specific window is employed to anticipate the next instant, CNN has outperformed the other three networks in the suggested study because it can capture sudden system changes. The benefits of creating a prediction model utilizing a hybrid network, which mixes two networks, have not been examined in this work.

[6] The strategy used in this study to forecast CNN stock prices is highly accurate and useful. Although the Stock Exchange of Thailand is the primary focus of this thesis, the established model can be used for any other stock exchange where a significant number of daily historical prices are available.

[7] Their findings demonstrate that the suggested model outperforms cutting-edge methods for time series forecasting.

[8] In comparison to CEEMD-LSTM and LSTM, the CEEMD-CNN-LSTM algorithm produced findings that were more accurate. The same was true for EMD-CNN-LSTM when compared to EMD-LSTM and LSTM. It suggests that including CNN in the suggested model was a good idea, and that using the CNN model in conjunction with LSTM could enhance the model's performance.

[9] This model outperforms other input permutations and SVM, providing a prediction accuracy of 87.86% in the remaining 10% of testing data after being trained on 90% of the complete data set technique by a minimum of 6%. The effectiveness of deep learning financial time series in the presence of significant noise is demonstrated in this paper.This strategy, which has been proved in this paper, can be used to anticipate financial markets with other factors and across whole different time periods.

[10] In this paper, the authors use the financial news articles dataset of 12 years collected from the Reuter website. Dataset features are extracted using Deep Learning Contextualized Word Representation (DCWR) approach and Independent Component Analysis (ICA) technique is used for feature reduction and HANet is their proposed algorithm for prediction of the stock market. The proposed model gives 92.5% accuracy.

[11] In this paper, the authors have collected daily historical data of NIFTY 50 index values for 6 years from Yahoo Finance website. They have derived certain parameters like high_norm, low_norm, and close_norm and fed them with stock variables in machine learning and deep learning models. The results show that deep learning models give far better performance than machine learning models and LSTM is the best among all, and univariate models perform better than multivariate models in LSTM based algorithms.

[12] This paper uses datasets from 2 stock markets, 50 company stock markets for Taiwan50 and 10 company stock markets for Indonesia. Candlestick chart images generated from computer graphics are used for prediction of stock market using the CNN model. The proposed model long-term trading days gives highest performance compared to other models for sensitivity, specificity, accuracy, and MCC. The authors also developed a website employing a proposed model for stock market prediction.

[13] Here the datasets are collected from different resources like NASDAQ, moneycontrol, nseindia website, finet, korea stock market. The papers discussed are divided on the basis of models used in it like Support Vector Machine, Neural Networks, Artificial Neural Networks, Convolutional Neural Networks, and so on. The various statistical parameters used by papers for measuring the performance are accuracy, root mean square error, mean square error, mean absolute error, mean absolute percentage error.

[14] This study tries to contradict the Efficient Market Hypothesis (EMH) and for this it uses and compares two deep learning models namely Long Short Term Memory Neural Networks (LSTM) and Feed Forward Multilayer Perceptron Neural Network (MLP). It uses 10 unique

stocks of the New York Stock Exchange and predicts their short-term prices of stocks. Papers concludes that MLP outperforms LSTM for short-term stock value prediction.

[15] This work compares different types of neural networks, Recurrent Neural Networks (RNN), Long Short Term Memory Neural Networks (LSTM), Back Propagation Neural Networks for stock market prediction using financial news sentiment prediction. The results show that when the model uses text data then it gives moderate accuracy but when used in conjunction with LSTM, the model accuracy vastly increases.

[16] In this paper, the authors use Chinese A-share market dataset for stock market prediction. The paper proposes a hybrid model that employees Convolutional Neural Networks (CNNs) for investors sentiment analysis and the result from CNN are integrated with technical indicators for stock market as the input for Long Short Term Memory Neural Network (LSTM) to predict the stock market value of one day ahead closing prices. The proposed model is compared with RNN, LSTM without sentiment data, SVM, CNN.

[17] In this paper, the authors have collected stock market data using Google Finance for the 30 DJIA Stocks for 15 years from Reuters website. VADER [HG14] and Google Sentiment API is used for the sentiment analysis, OpendIE extractor is used for converting sentence into tuple, and Word2Vec skip gram is used for converting words to the vector values. The proposed model of this paper uses LSTMs and specific convolutional architectures for stock market prediction.

[18] In this paper, the authors proposed an ensemble based neural network for stock market prediction using historical stock data and sentiment analysis. The data for sentiment analysis is collected from stocktwits microblog using python script prepared by SemEval-2017. 4 models are applied on the text data for getting hidden features, MLP feature driven model for extracting information from manually curated features, MLP simple word embedding for extracting information directly from the vector representation of a particular text, CNN for extracting local information, and last LSTM for extracting global information.

[19] In this paper, the authors have collected stock data using Twitter API (Twitter4J java's application) and used the opinion about the product produced by the company along with the tweets of the company. The sentiment values of 0,1,2 are manually assigned for the sentiment dataset instead of using different corpus dataset. They are using 2 techniques namely n-gram representation and word2vec representation. The algorithm used for stock market value prediction is Random Forest.

[20] In this paper,A survey on stock market was cloned,they researched nearly 138 articles from 2000-19.The stocks discussed in this paper are connected to health managing.It sectors all deep learning methods and analysis has made use to predict the stock more accurately.

## Material and methods

### Data Collection:

For sentiment analysis, we have collected stock tweets data of 5 industries instead of taking specific companies as taken in previous works done in this field. Industrial data is taken to introduce generalizability of our model and avoiding overfitting. We use Requests and BeautifulSoup library of python for scraping data from stocktwits website.

### Data Pre-processing:

Textual Dataset:

Pre-processing of textual data done so that we can remove unnecessary words, reduce the dimensions and convert it into vector form data that the sentiment classifier model can understand. For this firstly, we split the sentences of textual data into words, and after that remove the stop words to reduce the dimension such that the actual sentiment of the sentence is preserved and efficiency of model is increased. The remaining useful words are then transferred into vector representation for apply the deep learning model. We use, Word2vec tool on n words to get the word vector representation, it gives nxk matrix for n input words with k dimensions of the word vector.

Direct Stock Market Dataset:

We must divide our stock price data into a training set and a test set before we can create an LSTM model. Additionally, we will standardize our data so that all values fall between 0 and 1. In this case, training our LSTM model will simply require the closing prices from our dataset. As

| Normalization | Standardization |
| --- | --- |
| $X_{norm} = \frac{x - min(x)}{max(x) - min(x)}$ | $X_{stand} = \frac{x - mean(x)}{Std(x)}$ |

our training set, we will extract 80% of the closing prices from the stock data we have obtained. To prevent extensive computation, scale the stock prices between (0, 1) in the following step. Standardization and normalization are frequent techniques, as depicted in Figure. Normalization is advised, especially when working with RNN that has a Sigmoid function in the output layer. As already mentioned, we base our predictions on Open pricing. We only have one indicator or feature, specifically. However, using the same data processing techniques, we can add more indicators. To accomplish that, a new dimension for the amount of indicators must be added.

**Proposed Work:**

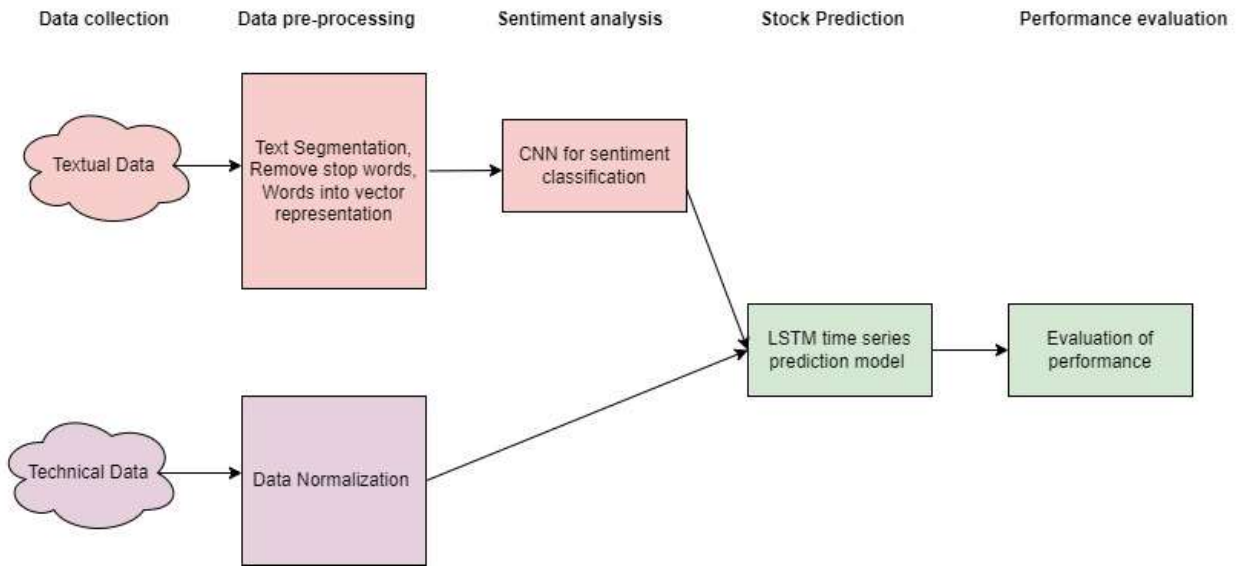



Fig. 1 Proposed system diagram

Deep Learning:

Deep learning is the subclass of machine learning that uses multiple hidden layers( unbounded number of layers of bounded size) to extract higher-level features from the input data. The algorithms that use deep learning concepts mainly use neural networks for computing large datasets and finding hidden patterns for predicting the results. The paper uses two algorithms: the Convolutional Neural Network model for text classification for sentiment analysis, the Long-Short Term Memory model for stock value prediction using the technical parameters, and the results from the sentiment analysis.

Convolutional Neural Networks for Sentiment Analysis:

A Convolutional Neural Network is a type of Artificial Neural Network that specializes in processing data with a matrix/grid-like structure. The paper uses CNNs for the classification of textual data. The CNNs models consist of the input layer, convolutional layer, max-pooling layer, and fully connected layer. The input layer consists of n words with k dimensions obtained from pre-processing step, the convolutional layer is used to perform convolution operation by sliding a convolution kernel with the size of h x k at the input layer to obtain a feature map c. One max pooling layer is used to extract the most important features. Finally, the fully connected layer applies a softmax classifier to obtain the classification result. As a result of the CNN model, the input words of text data are classified into negative and positive.

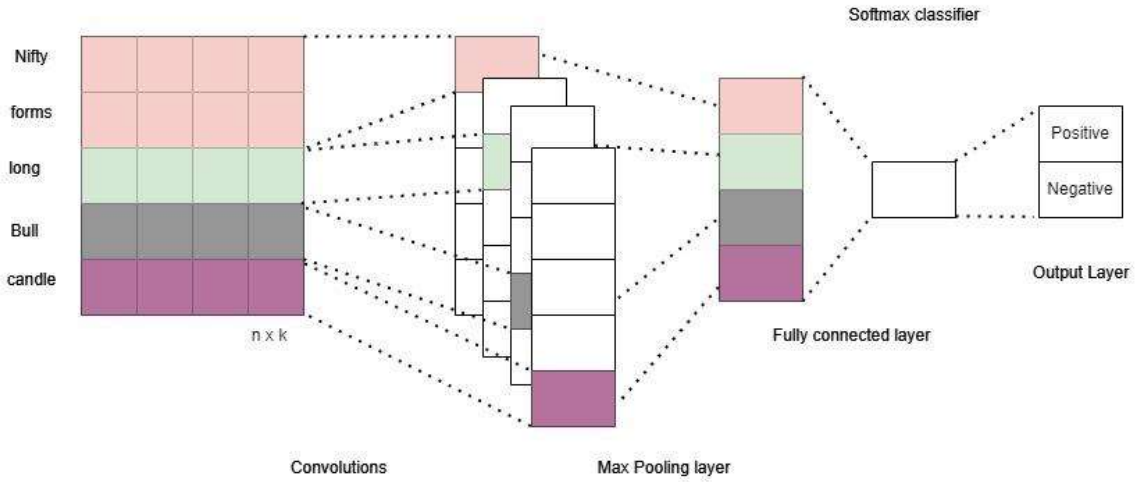



fig.2 : Structure of CNN model used.

Long Short-Term Memory Neural Networks (LSTMs) model for prediction:

Long Short Term Memory Neural Networks is a variety of Recurrent Neural Networks designed to avoid exploding and vanishing gradients for sequential data. LSTM has three gates: forget gate, input gate, output gate. Input gate can decide to put new information in the cell state, forget gate determines what information can be discarded from the previous cell state and what information can be retained for solving the gradient disappearance and is measured from zero to one, and finally, the output gate determines how much information should be exhibited.

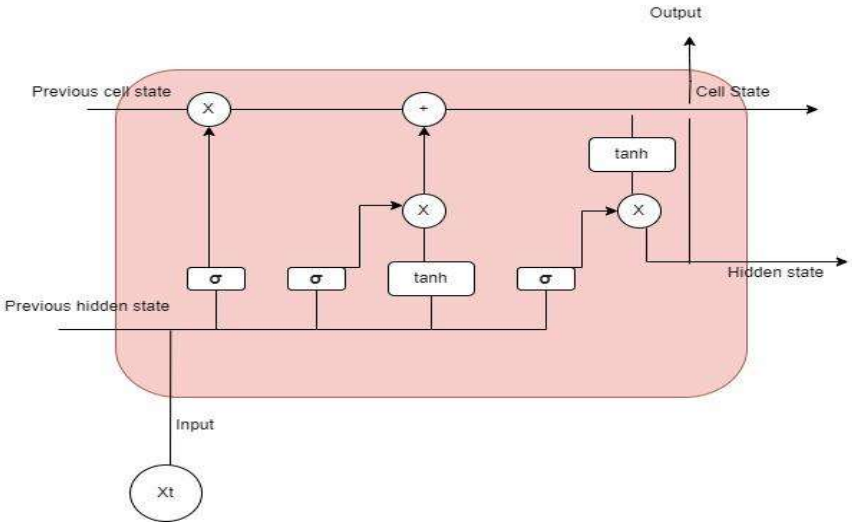
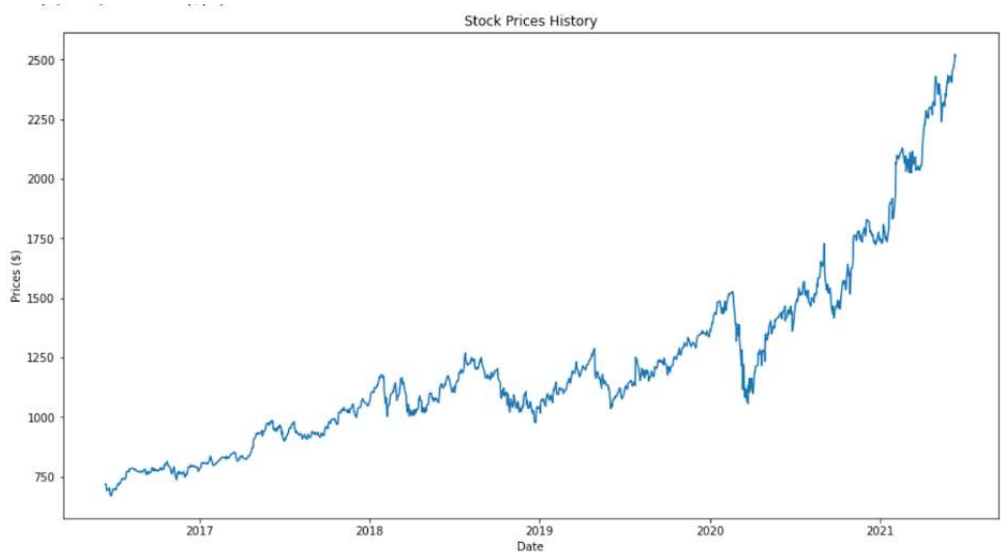



fig 3. Structure of a LSTM cell.

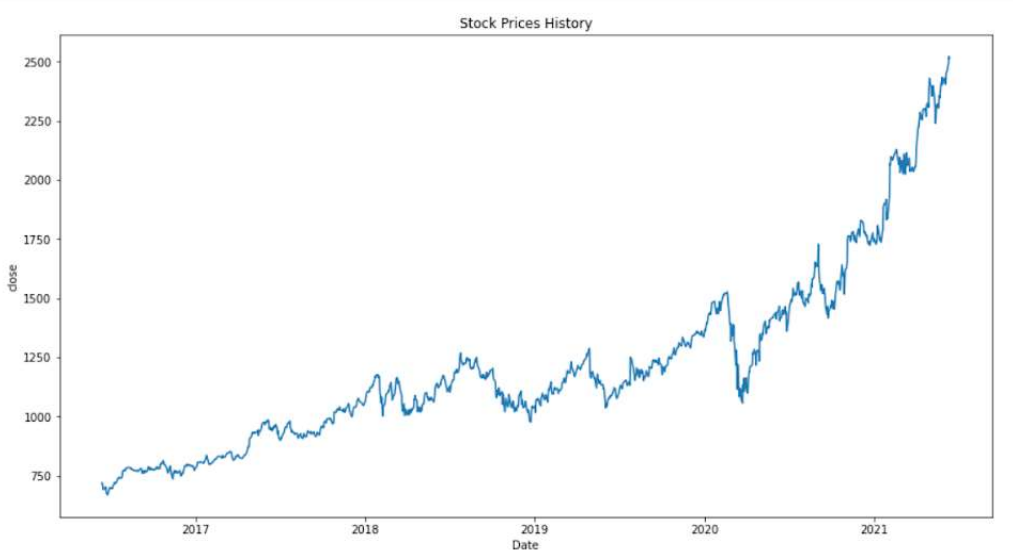
The LSTM model considers sentiment factors obtained from CNN sentiment analysis model and the technical indicators from historical stock market data as an input for making predictions, this input data is fed to the model to predict the one-day ahead closing price of the stocks. For LSTM, hyperparameter setting is important and performance of the proposed model under each hyperparameter setting is evaluated using Mean Absolute Percentage Error.

**Stock prediction based on LSTM Neural network:**

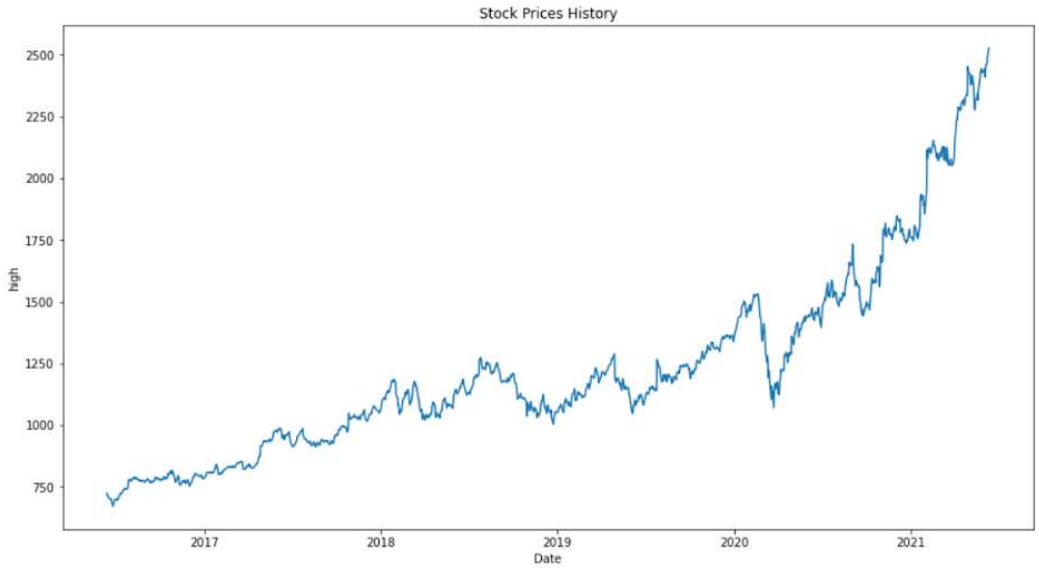
**Stock prices history:**



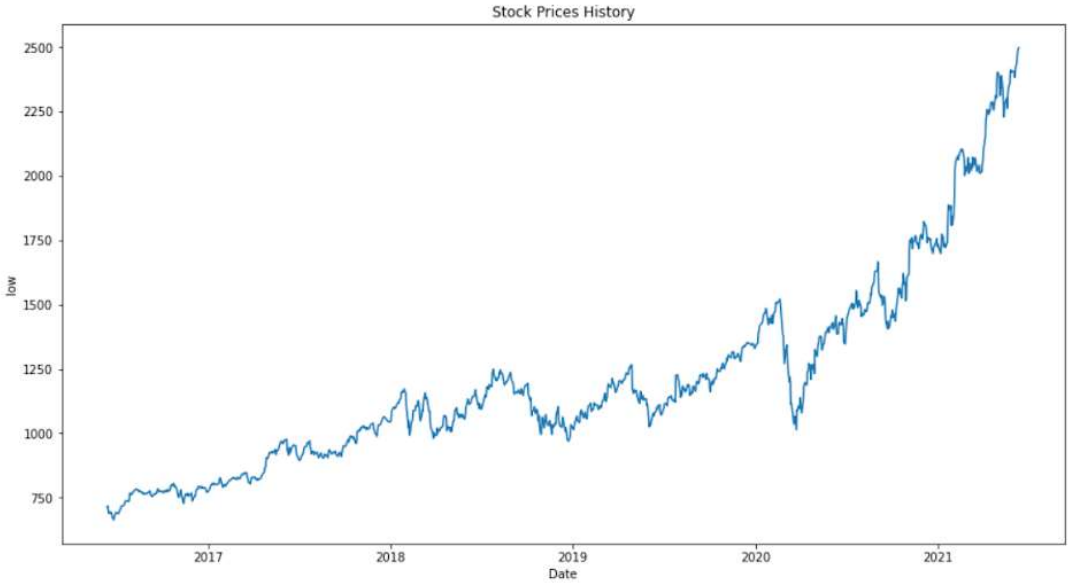

**Stock prices history based on close data:**



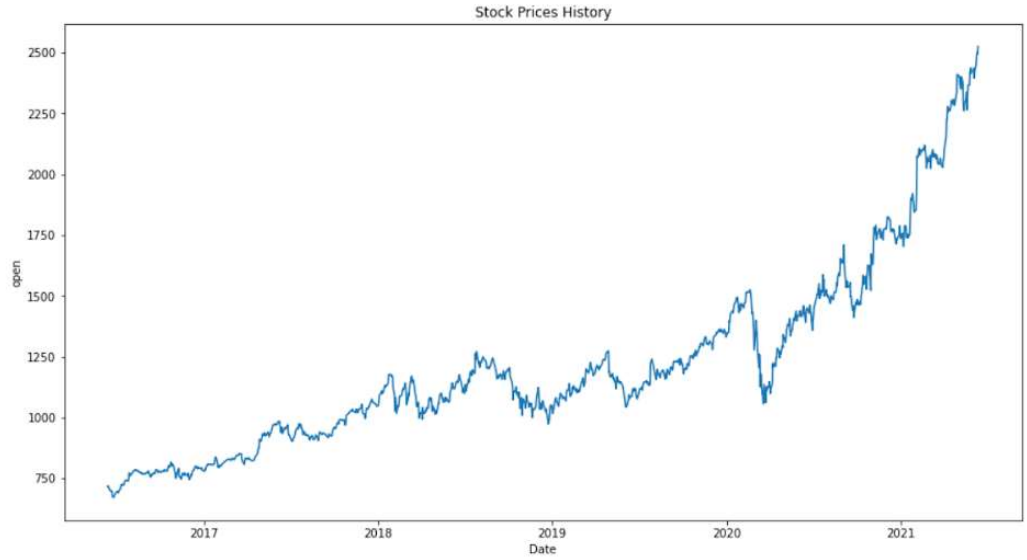
**Stock prices history based on high data:**



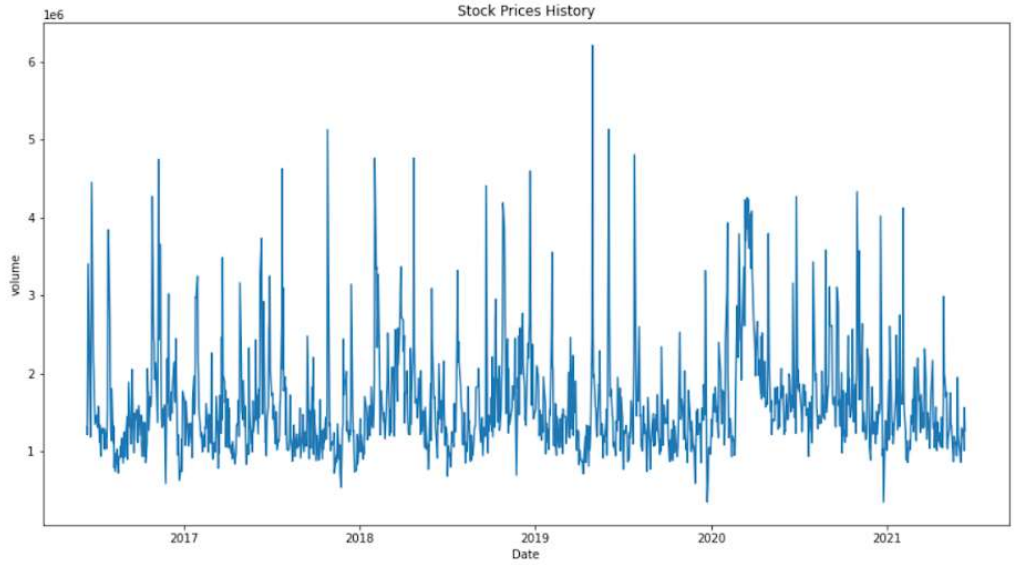

**Stock prices history based on low data:**



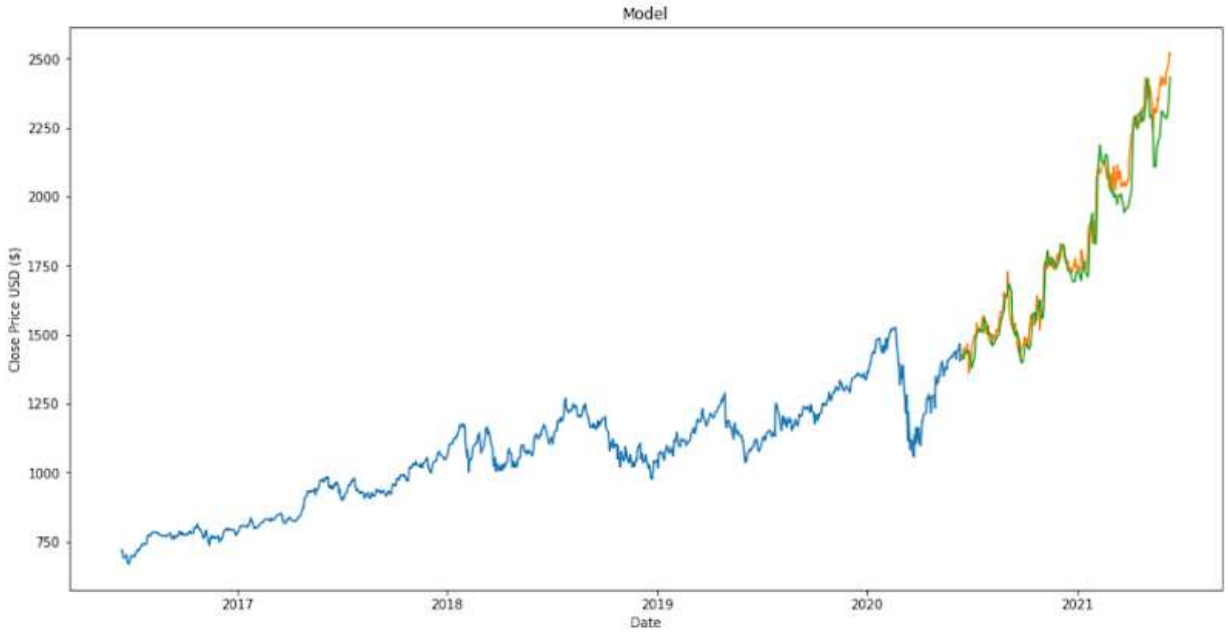
**Stock prices history based on open data:**




**Stock prices history based on volume data:**



**Result:**




From the resulting chart above, We can see how closely the predicted stock prices track the trend of the actual stock prices from the generated chart above. This demonstrates how well the LSTM performs when used with time series or sequential data, such as stock prices.Another excellent instrument for predicting stock price is LSTM. It's crucial to remember that the forecasted stock prices should not be relied upon as the only reliable indicator when choosing an investment strategy. This is due to the fact that the prediction is only based on historical price movements, which are frequently not the only factors influencing future price movements.

## Implementation

```
In [2]: df = pd.read_csv("stock_datasets/df_stocktwits_aapl.csv",
                         usecols=['date_time','tweet'])
        print(df.shape)
        df[:]
```

```
(190975, 2)
```

Out[2]:

| | date_time | tweet |
|---|---|---|
| 0 | 2011-11-15 | RT howardlindzon: Looks like Goldman $gs is t... |
| 1 | 2011-11-15 | $AAPL http://stks.co/17zI (Weekly Chart) Appro... |
| 2 | 2011-11-15 | $AAPL down -8.26% this morning? That is a real... |
| 3 | 2011-11-15 | RT Zguy: $AAPL down -8.26% this morning? That ... |
| 4 | 2011-11-15 | NEW POST: FROZEN TURKEYS http://stks.co/181s $... |
| ... | ... | ... |
| 190970 | 2023-01-04 | Dan Niles: In summary, my 2 overarching invest... |
| 190971 | 2023-01-04 | $AAPL |
| 190972 | 2023-01-04 | $AAPL$MSFT $GOOGL$AMZN\nI will buy more and ... |
| 190973 | 2023-01-04 | $AAPL bye apple, hello meta |
| 190974 | 2023-01-04 | $AAPL still don't have a reason to invest ... |

190975 rows × 2 columns

```
In [2]: df_aapl = pd.read_csv("stock_datasets/labeled_df_aapl.csv")
```

```
In [3]: df_aapl
```

Out[3]:

| | date_time | 0 |
|---|---|---|
| 0 | 2011-11-15 | -0.032980 |
| 1 | 2011-11-23 | -0.028850 |
| 2 | 2011-11-24 | 0.122656 |
| 3 | 2011-12-05 | 0.080717 |
| 4 | 2011-12-14 | -0.041253 |
| ... | ... | ... |
| 2691 | 2022-12-31 | 0.165750 |
| 2692 | 2023-01-01 | 0.132255 |
| 2693 | 2023-01-02 | -0.137685 |
| 2694 | 2023-01-03 | 0.036852 |
| 2695 | 2023-01-04 | -0.042010 |

2696 rows × 2 columns

```
dense (Dense)                (None, 5)                 645
=================================================================
Total params: 809,649
Trainable params: 809,649
Non-trainable params: 0
_________________________________________________________________
```

```
In [*]: model.compile(loss='sparse_categorical_crossentropy',optimizer='adam',metrics=['accuracy'])


        data_train = np.asarray(data_train)
        y_train = np.asarray(y_train)
        data_test = np.asarray(data_test)
        y_test = np.asarray(y_test)
        cnn_senti=model.fit(data_train,y_train,validation_data=(data_test,y_test),epochs=50,batch_size=100)
```

```
Train on 127953 samples, validate on 63022 samples
Epoch 1/50
127953/127953 [==============================] - 94s 734us/sample - loss: 0.0153 - accuracy: 0.4255
l_accuracy: 0.4279
Epoch 2/50
127953/127953 [==============================] - 89s 699us/sample - loss: 4.1925e-11 - accuracy: 0.4
```

**References**


1. Kaustubh Khare, Omkar Darekar, Prafull Gupta, Dr. V. Z. Attar: Short-Term Stock Price Prediction Using Deep Learning, IEEE, 2017.

2. Jordan Prosky, Andrew Tan, Xingyou Song, Michael Zhao: Sentiment Predictability for Stocks, ResearchGate, 2018.

3. Rosdyana Mangir Irwan Kusuma, Trang-Thi Ho, Wei-Chun Kao, Yu-Yen Ou, Kai-Lung Hua: Using Deep Learning Neural Networks and Candlestick Chart Representation to Predict Stock Market, preprint(ResearchGate), 2019.

4. Aditya Menon, Shivali Singh, Hardik Parekh: A Review of Stock Market Prediction Using Neural Networks, IEEE, 2019.

5. Louis Owen, Finny Oktariani: SENN: Stock Ensemble-based Neural Network for Stock Market Prediction using Historical Stock Data and Sentiment Analysis, 2020 International Conference on Data Science and Its Applications (ICoDSA).

6. Nan Jing, Zhao Wu, Hefei Wang: A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction, Elsevier, 2021.

7. Deepak Kumar, Pradeepta Kumar Sarangi, Rajit Verma: A systematic review of stock market prediction using machine learning and statistical techniques, Elsevier, 2021.

8. Sidra Mehtab, Jaydip Sen, Abhishek Dutta: Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models, Springer, 2021.

9. Saleh Albahli, Awais Awan, Tahira Nazir, Aun Irtaza, Ali Alkhalifah, Waleed Albattah: A deep learning method DCWR with HANet for stock market prediction using news articles, Springer, 2022.

10. Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda, Babita Majhi: Sentiment Analysis of Twitter Data for Predicting Stock Market Movements

11. Huy D. Huynh∗,L.Minh Dang,Duc Duong A New Model for Stock Price Movements Prediction Using Deep Neural Network

12. Pooja Mehta,Sharnil Pandya,Ketan Kotecha Harvesting social media sentiment analysis to enhance stock market prediction using deep learning

13. Mr. Tej Bahadur Shahi,Mr. Ashish Shrestha,Dr. Arjun Neupane,Dr. William Guo Stock Price Forecasting with Deep Learning

14. Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda, Babita Majhi Sentiment Analysis of Twitter Data for Predicting stock Market Movements

15. Jordan Prosky, Andrew Tan, Xingyou Song, Michael Zhao Sentiment Predictability for Stocks.

16. Louis Owen, Finny Oktariani SENN: Stock Ensemble-based Neural Network for Stock Market Prediction using Historical Stock Data and Sentiment Analysis, 2020 International Conference on Data Science and Its Applications(ICoDSA)

17. Mahindra Mahi Lagaha Kumbare,Christoph Lorhmann,Pasiluukka,Jari Porras Machine learning technique and data for stock market forecasting:A literature review,Elsevier(2022)

18. Parshv chhajer, Manan Shah, Ameya Khirasagar The applications of artificial neural networks,support vactor machines,and long short term memory for stock market prediction,Elsevier(2021)

19. Dingming Wu,Xialong Wung Jointly modeling transfer learning of industrial chain information and deep learning for stock prediction,Elsevier(2021)

20. Shourik Banik,Nonita Sharma,Monika Mangla,Sachi Nandan,Mohanty,Shithartth S LSTM Based decision supporting system for swing trading in stock market,Elsevier(2021)