CHAPTER 1

INTRODUCTION

MACHINE LEARNING

It is an AI application that, without explicit programming gives the ability to system to learn and improve with every experience. The main focus of ML is to develop programs which can be used by computers to access data and learn from the data. The first method in the process of learning is observing the data. Allowing the computers to learn automatically without any assistance from human and then allowing the computer to adjust accordingly after is the primary goal of machine learning.

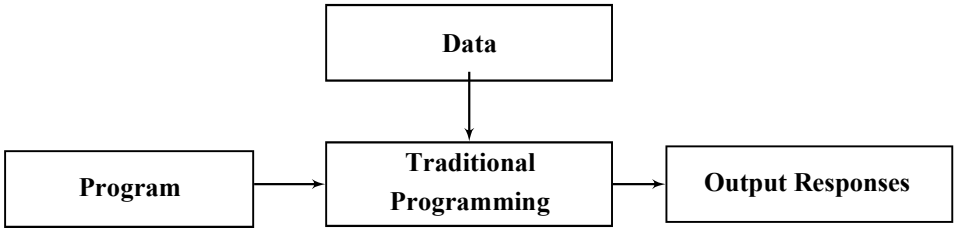

Figure 1.1: Traditional programming

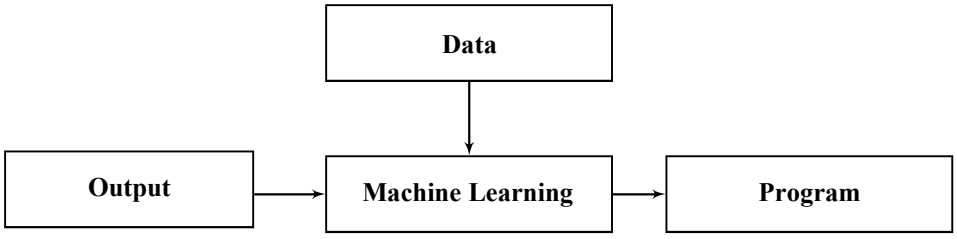

Figure 1.2: Machine Learning

Neural Networks

For IRIS we are using artificial neural network to forecast the future values of stock. Artificial neural networks (ANN) are determining systems that are virtually influenced

by the biological neural networks which constitute animal brains. These systems "learn" to perform tasks by examining examples, in general without being programmed with task-specific rules. For example, the image recognition system.

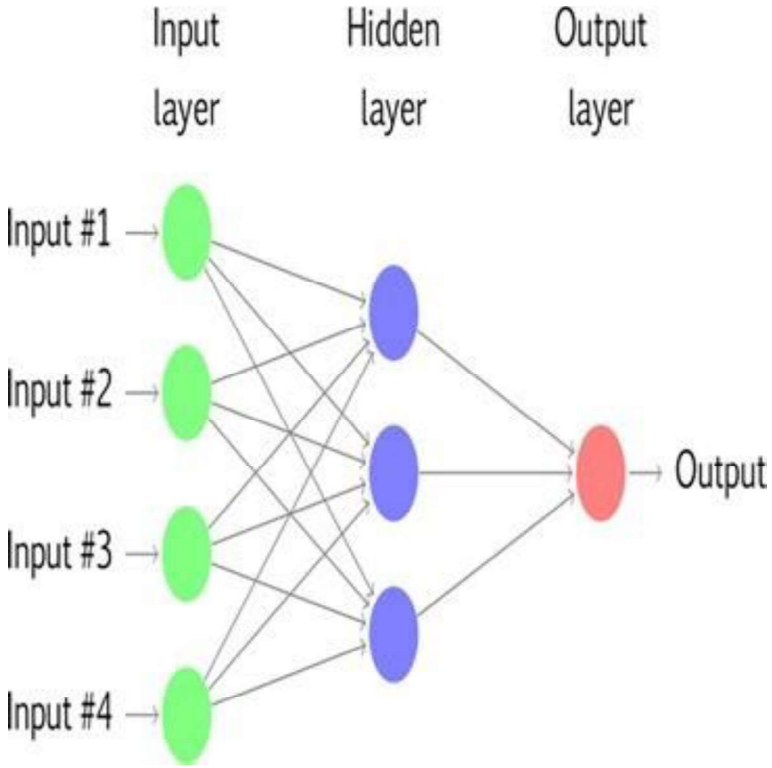

Figure 1.3: Neural Network

LONG TERM SHORT MEMORY

Long short-term memory (LSTM) is often used in the field of deep learning. It is considered as an artificial recurrent neural network (RNN) architecture. LSTM is different from standard feed-forward networks as it has feedback connections. It is not only limited to processing of single data points, as it can also process entire sequences of data.

For example, LSTM is applicable to tasks such as un-segmented, connected handwriting recognition, speech recognition and anomaly detection in network traffic or IDS's. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell.

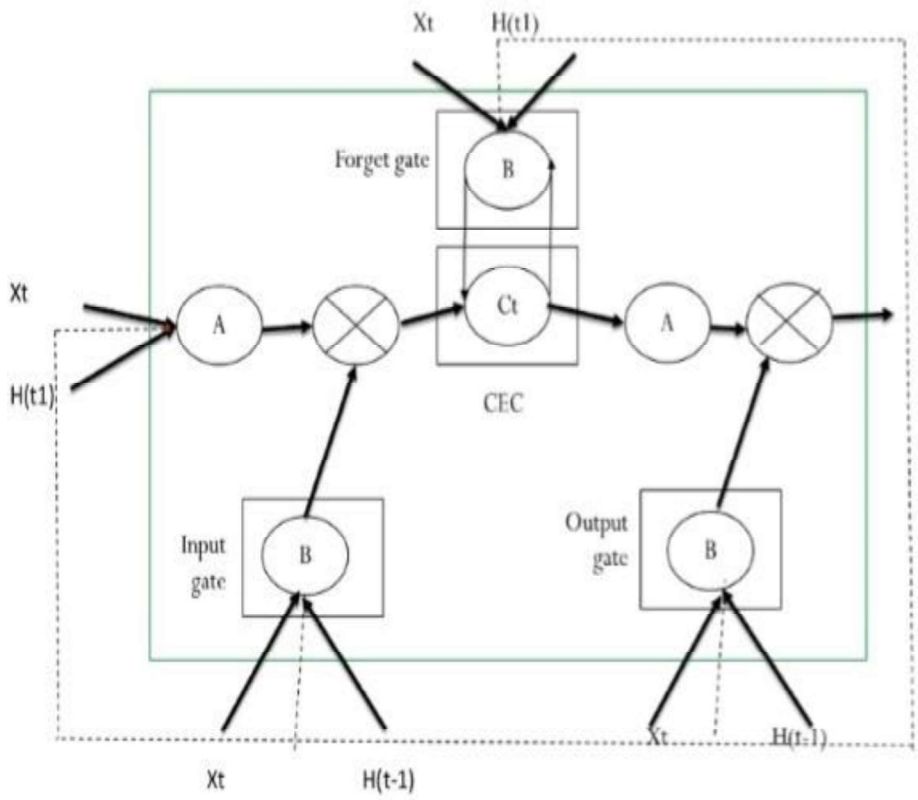

Figure 1.4: LSTM

Figure 1.4: LSTM

STOCKS

In a company, ownership share is represented by securities. These securities are stocks. A company can invest in their business and raise money to grow by issuing stocks Whereas Investors uses stocks to grow money and out space inflation.

Everyone who shares company's profit are called Shareholders. Stocks of a company are owned by them. First, Stocks are sold through a stock market exchange by public companies. These stocks are bought and sold by various investors among them-

selves via stockbrokers. A company's stock supply and demand that affects the price of the stock are tracked by the stock exchanges.

The stocks work on the following four prices:

i. Max- On a particular day, maximum stock price reached.

ii. Min- On a particular day, minimum stock price reached.

iii. Opening- Stock Price on the beginning of the day.

iv. Closing- Stock price on the end of the day.

For a particular stock of the upcoming days, Closing price will be predicted by IRIS.

CHAPTER 2

LITERATURE SURVEY

A Dual-Attention-Based Stock Price Trend Predic- tion Model With Dual Features

Author and year: Chen, Yingxuan, Weiwei Lin, and James Z. Wang. IEEE Access 7 (2019): 148047-148058.

Conclusion: The paper suggests a stock price trend predictive model (TPM) based on a decoder-encoder framework that predicting the stock price trending and its duration adaptively. The predicted model consists of two phases; first, an extraction method with dual-feature based on different time lengths is proposed to get more information related trends from the market data. The traditional methods only used to extract features from information at some particular time period, but this proposed model applies the Prime Lending Rate (PLR) method and Convolutional Neural Network (CNN) to draw out the long-term time-related features and the short-term structural features from the market data. In the second phase of the proposed trend predictive model (TPM), an encoder-decoder framework based on dual-attention is used to find and integrate relevant dual features and predict the trends in stock prices.

Which artificial intelligence algorithm better pre- dicts the stock market?

Authors and year: Chen, Lin, IEEE Access 6 (2018): 48625-48633.

conclusion: It is difficult to predict stock index futures because of the unpredictable nature of stock market factors. There is a long history of efforts to develop an efficient

method to predict stock index futures. Due to recent increase in use of AI neural networks, success in nonlinear approximation have increased. Without previous predictive information, features can be extracted from a big data environment while studying financial market. In this paper it is proposed to compare various neural networks and after comparing the results, selecting the best algorithm according to the data which is being used for prediction.

Stock Volatility Prediction by Hybrid Neural Net- work

Authors and year: Wang, Yujie, IEEE Access 7 (2019): 154524-154534. Conclusion:

To reduce risks in investments, the most important role is played by Stock Volatility Prediction. It is a subject which is undergoing intense study when we consider prediction using time series. Nevertheless, the movement of stock price has certain other factors as well other than its historical trend. It also depends upon some correlated social factors. The above paper suggests a hybrid time-series predictive neural network (HTPNN) that considers the affection of news. To make the model more efficient by sparse automatic encoders, the attributes of news headlines are demonstrated as distributed word vectors and are dimensionally reduced.

Stock market prediction via multi-source multiple instance learning

Authors and year: Zhang, Xi, IEEE Access 6 (2018): 50720-50728. Conclusion:

Extraction of effective indicators like sentiments and events etc. are done by researchers from web as the information grows facilitates in prediction. Stock market movement is affected as indicators is obtained from data sources that does uncover the factors. Consistencies obtained from different data sources are exploited and a multiple instance multi-source instance model is developed which further forms a comprehen-

sive framework which combines events and sentiments along with the quantitative data

Forward forecast of stock price using sliding-window meta heuristic-optimized machine-learning regres- sion

Authors and year : Chou, Jui-Sheng, and Thi-Kha Nguyen, IEEE Transactions on Industrial Informatics 14.7 (2018): 3132-3142.

In a ever changing environment of the stock market, the non-linearity of the time series is very noticeable, rapidly affecting the stock price predictions. Thus, using sliding-window metaheuristic optimization this paper proposes an intelligent time series calculation system. It can help home brokers who do not know how to invest in after studying the stock market properly.

CHAPTER 3

SYSTEM ANALYSIS

OBJECTIVE

An investor needs to be aware about the stock market, before making any investments. As making good investment at a bad time can lead to a huge loss. Hence, the change or variations in the prices of stock in the stock market depends upon the sentiments of many investors. Prediction of stock market should have the potential to predict the consequence of current events on the investors who are investing in stocks. There are many events which can directly or indirectly effect the price of stocks. Some of these events are political leader statements, news on scams and other similar things. Some events which are taking place internationally such as sharp movements in commodities and currencies etc also have its effect on the stock market. These changes have an impact on the earnings of the corporate, which can affect the investor's sentiments. These variables are hard to predict by the investors correctly and consistently. These are some of the challenges that make prediction of stock price difficult. Hence, the main challenge is live testing, because the new algorithms having a lot of factors like price variations, and unwanted news and noise continuing to overflow the markets every day, comparison of the efficiency and validity of these algorithms brings up another dispute for the developers. Another compelling aspect is its self-destructing nature. In other words, if an algorithm can use an uncommon approach to generate high profits, then distributing it to the market participants will leave the uncommon approach of no use. A lot of unrequired attention is being given to the viewpoint analysis based on data retrieved from Twitter or news due to the increasing effect of social media on different aspects of our lives. Numerous fake news is being posted on the internet by multiple unreliable and unverified sources, hence social media data can be untrustworthy and tough to process

Expected Impacts on Academics/Industry

1. It uncovers the future market behaviour which always helps the investors to understand when and what stocks can be purchased for the growth of their investment.
2. As the understanding of how market moves increase, Investors will be able to avoid financial crises as they are better equipped.
3. Existing strategies should be evaluated from a rigorous scientific perspective to provide new strategies with quantitative evaluation.
4. It allows simple running of the results based on the latest data that is fed-in or directly extracted from sources without any human bias.
5. It allows trend analysis within a matter of minutes in your hand and hence improves decision making.
6. It eliminates the need of an advisor and hence reducing another human factor and hence improving the result.

CHAPTER 4

SYSTEM DESIGN

Objective

To introduce a system that will predict the next closing value of the stocks of a particular company. The system will use the data extracted from yahoo to train itself and predict the next closing value of the stocks based on the trends it is trained with.

The user input form will have 3 input fields in total, as follows:

1. Name of the stocks- The stocks name of the company whose price prediction is to be conducted. For example, APPL for Apple.
2. Starting date- The date from which the user wants the system to train itself to find out the trends in the price variations.
3. Today's date- Today's date is asked to predict the next closing price by the system.

The developed model will fetch the data right from the starting date till todays date to train itself according to the variations in the stock prices since then. System uses LSTM algorithm to train for a given data.

USE CASE DIAGRAM

There are two users in the entire system. First is the end user. The end user inputs the required information into the system i.e. the name of the company, starting date and today's date. And finally view the predicted result that is, the predicted closing value of the stock. Second is the server. The server collects data right from the starting date and then train itself to produce the predicted next closing price of the stock. The server then finally updates the system with the predicted next closing value of the stocks.

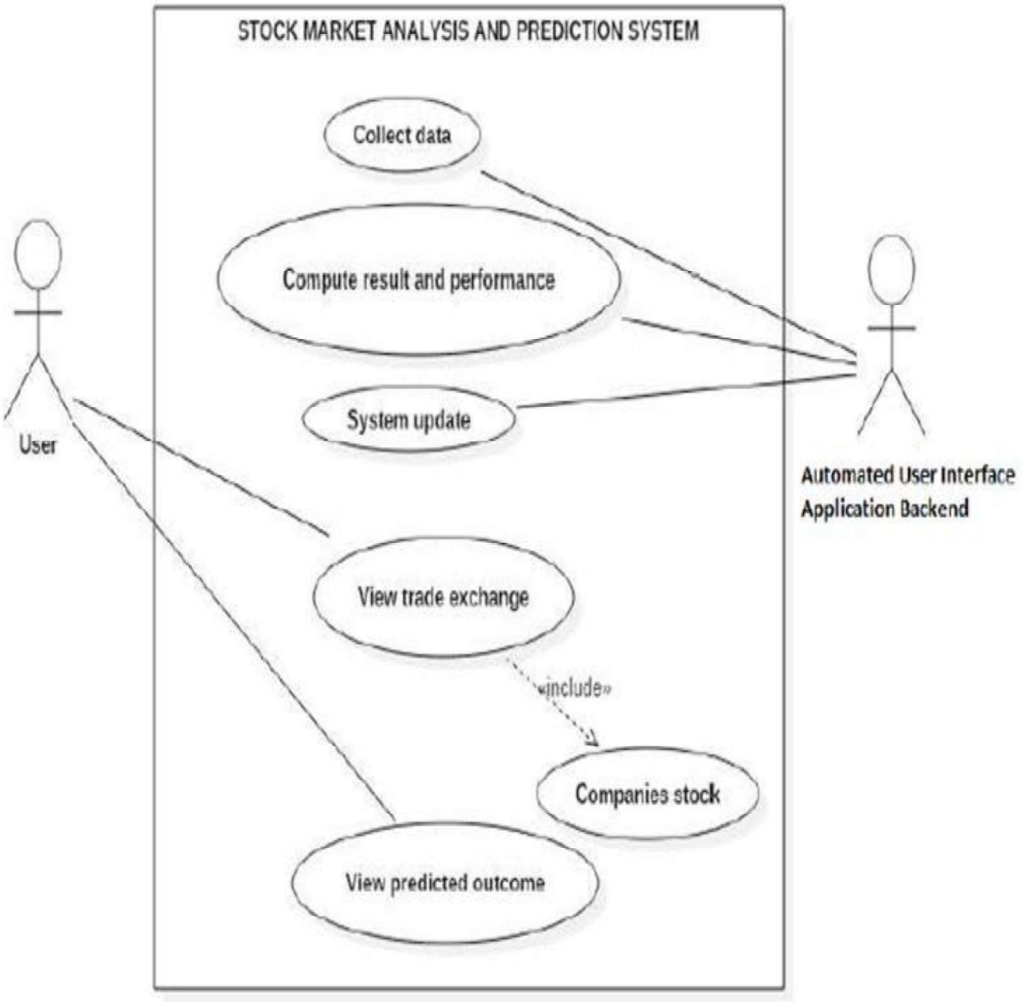

Figure 4.1: USE CASE DIAGRAM

ER DIAGRAM

The Entity-Relationship diagram depicts the relation between various entities of the system. The 3 entities are: the Market, the admin/user and the prediction model. The admin or user monitors the market and its various attributes. The market attributes are date, high price, low price, volume, open price, etc. The user then utilise the knowledge gained to predict the stock value according to his need. The Prediction model utilises the user inputs and data fetched to predict the next closing value of the stock which is then observed by the user of the system.

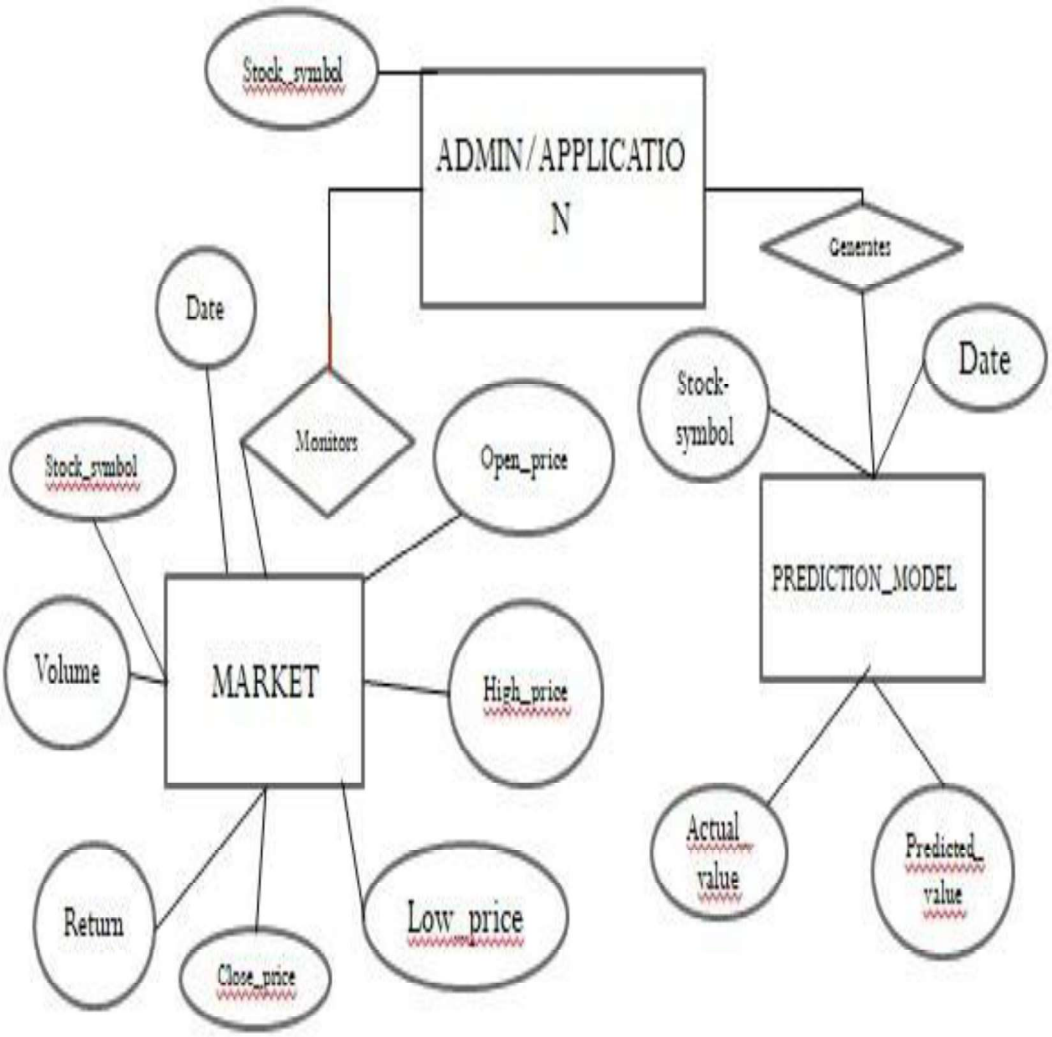

Figure 4.2: ER DIAGRAM

DATA FLOW DIAGRAM

The data flow diagram depicts the flow of the processes taking place in the system during an evaluation. The user inputs the company and other details. The system then fetches the data from the data source. The system then plots data and trains the model according to the trends obtained from the data fetched. The system then uses the trained data to predict the next closing value of the stocks and a result is obtained.

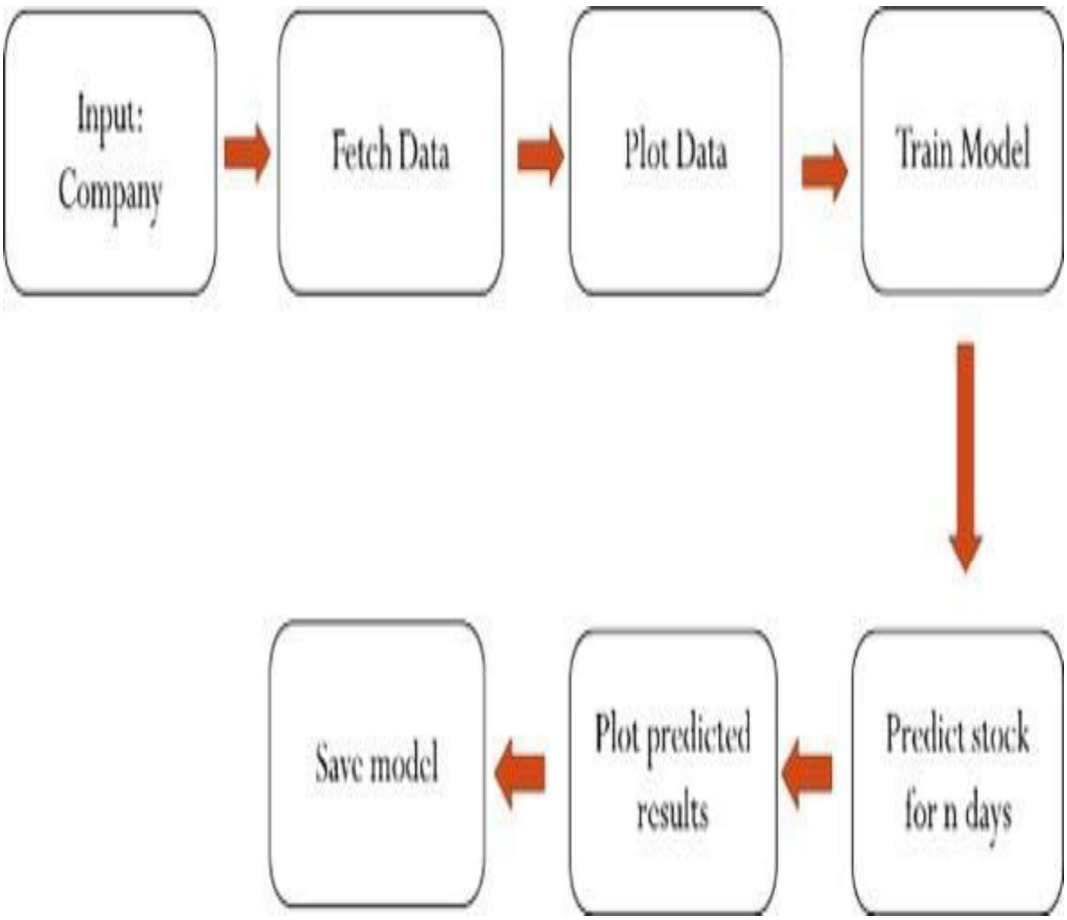

Figure 4.3: DATA FLOW DIAGRAM

STOCK RATES OF THE COMPANY

The stock rates of the company are stored in a tabular form in the system as shown below:

Date	High	Low	Open	Close	Volume	Adj Close
2012-01-03	58.928570	58.428570	58.485714	58.747143	75555200.0	50.994907
2012-01-04	59.240002	58.468571	58.571430	59.062859	65005500.0	51.268970
2012-01-05	59.792858	58.952858	59.278572	59.718571	67817400.0	51.838169
2012-01-06	60.392857	59.888573	59.967144	60.342857	79573200.0	52.380054
2012-01-09	61.107143	60.192856	60.785713	60.247143	98506100.0	52.296970
...	...	...	...	...	...	...
2020-04-13	273.700012	265.829987	268.309998	273.250000	32755700.0	273.250000
2020-04-14	288.250000	278.049988	280.000000	287.049988	48748700.0	287.049988
2020-04-15	286.329987	280.630005	282.399994	284.429993	32788600.0	284.429993
2020-04-16	288.200012	282.350006	287.380005	286.690002	38845800.0	286.690002
2020-04-17	286.950012	276.859985	284.690002	282.799988	53756600.0	282.799988

Proposed Algorithm

1. df= Read$_{d}$*ata*(*source, comp$_{n}$ame, start$_{d}$ate, present$_{d}$ate*)
2. *filter$_{c}$olumn* = (*stockpricetobepredicted*)
3. *scale*(*filter$_{c}$olumn*)
4. *train$_{l}$en* = $0.8 * filter_{c}olumn$

```
5.traindata = filter[0 :trainlen]
6.forarange(60, len(trainlen)):
atrain.append(traindata[a − 60 : a, 0])
btrain.append(traindata[a, 0])
7.atrain, btrain =np.array(atrain), np.array(btrain)
8.atrain = np.reshape(atrain, (atrain.shape[0], atrain.shape[1], 1))
9.Model = Sequential()
10.Model.add(LSTM(50, returnsequences = True, inputshape = (atrain.shape[1], 1)))
11.Model.add(LSTM(50, returnsequences = False))
12.Model.add(Dense(25))
13.Model.add(Dense(1))
14.Model.compile(optimizer =' adam', loss =' meansquarederror')
15.Model.fit(atrain, btrain, batchSize = 1, epochs = 1)
16.testdata = filtercolumn[traininglen − 60 :, :]
17.Y test =dataset[traininglen :, :]
18.foriinrange(60, len(testData)) : atest.append(testdata[i − 60 : i, 0])
19.atest = np.array(atest)
20.atest = np.reshape(atest, (atest.shape[0], atest.shape[1], 1))
21.predictions = Model.predict(atest)
22.predictions = scaler.inversetransform(predictions) 23.newdf
= Readdata(source, compname, startdate, presentdate) 24.newdf =
compquote.filtercolumn
25.last60Days = newdf [−60 :].values
26.last60DaysScaled = scaler.transform(last60Days)
27.Atest.append(last60DaysScaled)
28.Atest = np.array(Atest)
29.Atest = np.reshape(Atest, (Atest.shape[0], Atest.shape[1], 1))
30.predictPrice = model.predict(Atest)
31.predictPrice = scaler.inversetransform(predprice)
32.print(predictPrice)
```