Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

Stock Market Analysis using LSTM in Deep Learning

D. Mahendra Reddy
Assistant Professor (Ad-hoc),
Department of Computer Science and Engineering;
JNTUA College of Engineering; Pulivendula,
Andhra Pradesh, India.

K. Ashok Kumar Reddy
Department of Computer Science and Engineering;
JNTUA College of Engineering; Pulivendula,
Andhra Pradesh, India.

H. Veeresh Babu
Department of Computer Science and Engineering;
JNTUA College of Engineering; Pulivendula,
Andhra Pradesh, India.

Y. Saileela
Department of Computer Science and Engineering;
JNTUA College of Engineering; Pulivendula,
Andhra Pradesh, India.

Abstract: **Stock market prediction is that the scene of trying to complete the long-run value of company stock. Analysis of the stock price we take the price. By using the neural network, we develop a model. within the neural network, we use a recurrent neural network that remembers each and each information through time. LSTM networks maintain to stay contextual information of inputs by associate a loop that permits information to travel from one step to the following. These loops make recurrent neural networks seem magical. Each x train has the last 60 days value for the present-day y data. this can add more accurate results when a measure to existing stock price prediction algorithms. The network is trained and evaluated for accuracy with various sizes of knowledge, and also the results are tabulated.**

Keywords: ***Stock Market, LSTM, RNN (Recurrent Neural Network).***

I. INTRODUCTION

The stock price of a company is changed every day based on their exchange of products and raw goods. A company provides the shares to increase the production of their company production. To predict the value of the share is not an essay because it will change based on many factors. Many investors try to invest their amount in a company to get profit. Here we have a problem with finding the company share value is going to increase or decrease for that day end. In this case, we are trying to use technology for attaining perfect value. Technology is evolving swiftly. There are many feathers to be acknowledged while we try to build a machine learning model. In a day stock market has three major values high, low, open and close. Scrutinize the historical data to predict value is going to increase or decrease. Many machine learning algorithms are used to predict the value but as these types of time forecasting problems, we can apply SVM and Recurrent Neural Network. SVM and LSTM have their unique functionality for time series forecasting try to implement them and try to increase the accuracy. By using LSTM we create a model and from the past data, we train that model and predict 7 companies and analyze the data from that results. In some cases it difficult to predict the value because due to other feathers the value goes high or low for the day. The cases the value goes wrong are when the company tries to implement new things or any political effects generally called a new feather. In the model, we just included only past stock datasets and forms that we try to recognize the patterns in the data sequences.

A. Existing System

In the existing system, SVM and Backpropagation Algorithm there is no which won't do dropout process. Because of this, unwanted data have been processed which leads to wastage of time and memory space. The prediction of future stock price by SVM and Backpropagation Algorithm is less efficient because of processing unwanted data. The SVM and Backpropagation Algorithm which is used in the existing system is not that effective in handling non-linear data. So, in our proposed future stock price prediction is done using LSTM (Long Short Term Memory) which is a higher accurate value for the next day than SVM and Backpropagation Algorithm.

B. Proposed System

In the proposed system we try to find the accurate value of the next day closing value that helps the investors to invest or sell their shares. Long Short Term Memory (LSTM) is an artificial neural network in the field of deep learning. LSTM is an advance Neural network with having a memory cell that stores a small amount of data for further references. LSTM has feedback links that make it a "general-purpose computer". LSTM can also process an entire series of data not only single value like image. Because of the dropout process which takes place in the LSTM algorithm, it is comparatively faster than SVM and Backpropagation. LSTM algorithm is more suitable in predicting the future stock price than the SVM and Backpropagation algorithm because of removing the undesired data. The time and memory consumption are also reduced when compared to the exciting system due to the dropout process. LSTM algorithm is more proper in handling non-linear data. We predict the 10-company stock price and store them in a tabular format and visualize them.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

II. SYSTEM ARCHITECTURE

A. *Download Share price Data*

There are many sources for getting the share price data. Python has a library to get the data from the internet. Pandas_datareader has 4 parameters that will indicate how and where to get the data to the API. The first parameter is it specify the stock name of a company. next, it has a data source parameter that indicates that from where the API should you want to collect the data and the next parameters are starting date and ending date. There are many API to collect the data from the internet every API has some limitation of requesting data. so, by using Pandas_datareader I collected the 7 companies. the data files are in CSV format that can be used easily in Jupiter notebook. data set has Date, High, Low, Open, Close, Volume, Adj Close.

```
import pandas_datareader as web
from datetime import datetime
companys = ['GOOGL','MSFT','AAPL']
for i in companys:
    df=web.DataReader(i,
                      data_source='yahoo',
                      start='1990-01-01',
                      end=datetime.date(
                          datetime.now()))
    df.to_csv(i+".csv")
```

Fig. 1. Collecting data using Pandas_dataread.

B. *Data Set*

As we know the data set is the starting point for everything it should have full-fledged data to make the machine learn about the problem. Datasets can be generated or developed from the scrap information available on the internet. Some problems we have to create a dataset that makes sense that tells how to respond based on real-time inputs for the problem datasets can be gathered from the internet every day. A dataset is a collection of data. Most commonly a data set has contents of a single database table, or a single statistical data matrix, where every column of the table describes a particular variable, and each row matches a given member of the data set in question. The data set lists values for each of the variables, such as the height and weight of an object, for each member of the data set. Each value is recognized as a datum. As we know the data set is the starting point for everything it should have full-fledged data to make the machine learn about the problem. Datasets can be generated or developed from the scrap information available on the internet. Some problems we have to create a dataset that makes sense that tells how to respond based on real-time inputs for the problem datasets can be gathered from the internet every day. Here we keep all our data in the form of CSV files. In computing, a comma-separated values (CSV) file may be a document that uses a comma to separate values. CSV file stores tabular data (numbers and text) in plain text. Each line of the file may be a data record. Each record consists of 1 or more fields, separated by commas. The use of the comma as a field separator is that the source of the name for this file format. Our dataset is kept in tabular format in CSV with values such as date, open, high, low, last, low, total trade and turnover values.

Date	High	Low	Open	Close	Volume	Adj Close
19-08-2004	52.08208	48.02803	50.05005	50.22022	44659000	50.22022
20-08-2004	54.59459	50.3003	50.55556	54.20921	22834300	54.20921
23-08-2004	56.79679	54.57958	55.43043	54.75475	18256100	54.75475
24-08-2004	55.85585	51.83684	55.67567	52.48749	15247300	52.48749
25-08-2004	54.05405	51.99199	52.53253	53.05305	9188600	53.05305
26-08-2004	54.02903	52.38238	52.52753	54.00901	7094800	54.00901
27-08-2004	54.36436	52.8979	54.1041	53.12813	6211700	53.12813
30-08-2004	52.7978	51.05606	52.69269	51.05606	5196700	51.05606
31-08-2004	51.90691	51.13113	51.2012	51.23624	4917800	51.23624
01-09-2004	51.53654	49.88488	51.4014	50.17517	9138200	50.17517
02-09-2004	51.23624	49.51952	49.64465	50.80581	15118600	50.80581
03-09-2004	50.92092	49.70971	50.52552	50.05505	5152400	50.05505
07-09-2004	51.05105	49.85485	50.55556	50.84084	5847500	50.84084
08-09-2004	51.56657	50.3003	50.42042	51.2012	4985600	51.2012
09-09-2004	51.40641	50.55055	51.31631	51.20621	4061700	51.20621
10-09-2004	53.33333	50.7007	50.85085	52.71772	8698800	52.71772

Fig. 2. Dataset of GOOGL.

C. *Data Preprocessing*

Data set is a collection of feathers and N number of rows there are many values and values are in different formats. In a dataset, they may be duplicate values or null values that may lead to some loss inaccuracy and there may be dependent. Data have been collected from different sources so there use a different type of format to notate a single value like gender someone represents M/F or Male/Female. The machine can understand only 0 and 1 so an image will be in 3-dimension data should be reduced to a 2-dimension format like data show to free from noisy data, null values, an incorrect format. Data cleaning can be performed by panda's tabular data and OpenCV for images.

D. *Feather Scaling*

The dataset consists of numerical value which are constantly increasing day by day in the data set the closing value gradually increasing from 1 to 2000. As in the neural network target variable with a large spread of values, in turn, may result in large error gradient values causing weight values to change dramatically, making the learning process unstable. So, we scale the dataset into a small scale of (0,1) or (0,2) that make simple and reduce the loss. By scaling it easy for the model to store similar data for further reference in the further. We scale all the data to 0 to 1 and then try the model and pass the scaled value to get the output and reconvert the scale to an original scale value.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

```
[0.09975669]
[0.07931873]
[0.06457421]
[0.06944039]
[0.07445255]
[0.06900244]
[0.05406326]
[0.05090025]
[0.05776156]
[0.04861314]
[0.04462287]
[0.03961071]
[0.04418492]
[0.04175183]
[0.04559611]
[0.04515816]
[0.04705596]
[0.04642336]
[0.05138687]
```

Fig. 3. Closed Values of the GOOGl after Scaling.

E. LSTM Training and Evaluation

Long Short Term Memory is an Artificial neural network that has a feedback connection that makes the network work efficiently. LSTM has the advantage of handling sequential time series data not only a single valid data. It has three cells that regulate the functioning, input, and output of the network. This type of network is suitable for classifying and predicting the time series problems which we can make things better untestable to a network.

There are several architectures for LSTM but most commonly used are three cell ones that make the network work efficiently and predict the values accurately. An **input gate**, an **output gate** and a **forget gate**.

The **Input Gate** is the point where we pass the data into the network that we want to teach the model. Here there is another input from the output gate that is also called as a feedback signal that is the main one that was model learn what mistake it hades done and learn how to overcome from that loss. In these input gate, we use the sigmoid function and tanh function to combine the input and hidden value and get the output in the range -1 to 1.

The **Output Gate** is the last gate in the circuit we it gives the output value and it has an important duty that has to decide what data should be hidden for the next time. First, the previse hidden data and input passed to sigmoid function then pass the input value and sigmoid output multiply at tanh function that decides what should be stored for the next step.

The **Forget Gate** is the main gate that stores the previous information for farther references. The input from the previous data and the current data to the sigmoid function that gives the value 0 or 1. If we get 1 the forget gate store the data otherwise it will forget the present data.

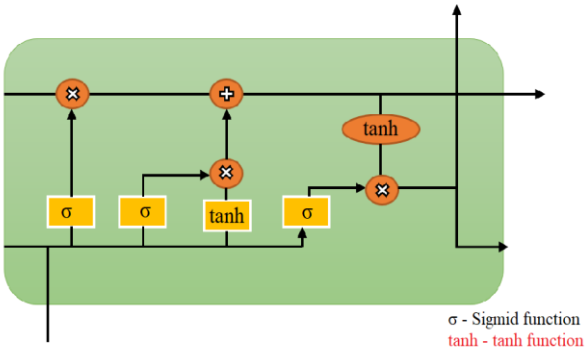

Fig. 4. LSTM architecture diagram.

III. PROPOSED WORK

By using LSTM, we try to implement the predicter model for the top 10 companies. For that, we have done some data modification that made the algorithm more efficient to find the accurate value. First, we get the data from the yahoo server by using the pandas web Reader that uses four parameters name of the company, server name, starting date and ending date.

That data has been scaled because we have ranged value from 0 to different so for making the model more efficient, we scale the input data from 0 to 1. For that, we have a package sklearn that has a scaling method to convert the input data to a range from 0 to 1. The main point is to create an x_train and y_train after the scaling we create training data we take 60 days of previous data for 'y' day value. So, we create train data for every 'y' day that the previous 60 days values are responsible for that day. Like that we create a training data set.

We try to create a network model with two LSTM layers and 2 dense layers that make the model try to find the closing value for the next day. Sequential methods are used because of the advantages of finding patterns of similarities. it helps in finding the next event for that particular X data. Model Compile defines the loss function, the optimizer, and the metrics. That's all. It has nothing to do with the weights and you can compile a model as many times as you want without causing any problem to pretrained weights. LSTM. The compiler has optimizer Adam that makes the network learn the value. We use the loss as have a mean squared error than try to reduce the loss that has accrued at the time of learning.

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 60, 50)	10400
lstm_2 (LSTM)	(None, 50)	20200
dense_1 (Dense)	(None, 25)	1275
dense_2 (Dense)	(None, 1)	26

Total params: 31,901
Trainable params: 31,901
Non-trainable params: 0

Fig. 5. LSTM Model.

Then we try to fix the x train data and y train data to the model fit to make the model learn from the historic data. The

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

epochs are used to make the model learn the same data repeated times. At epochs 2 is apply for that we got the nearest value for the next day closing value. Batch size is 1 because each value is individual and it is independent that makes the prediction accuracy.

Then we create x train and y train to predict the accuracy of the model and predict the values for x train and get the y to predict value. We compare both y train and y predict values. The RMSE is 5.77 that gives the assurance of that we can find the max closest value for the next day closing value.

IV. RESULT AND ANALYSIS

To find out the next day's closing value we create one new x list of 60 days values from today to past 60 days value that has to be scaled and passed to the LSTM trained model to predict the value that will give you the scaled value for the next day. We inverse the scaled value to get the original value.

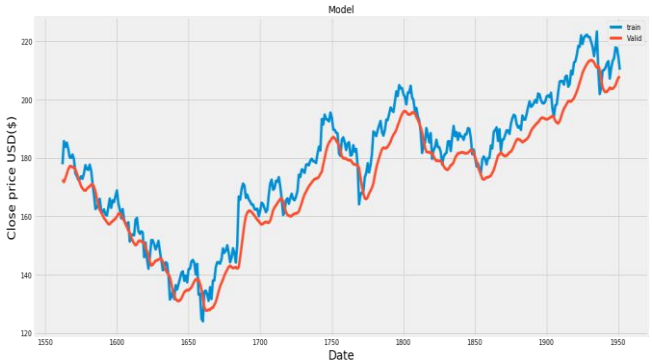

Fig. 6. Visualizing the predicted GOOGL price.

Like that we have find 7 company stock closing price for 10 days and visualize the data.

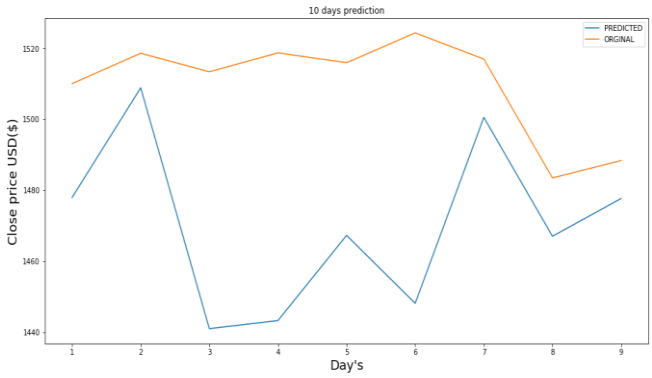

Fig. 7. Visualizing the GOOGL 10 days prediction.

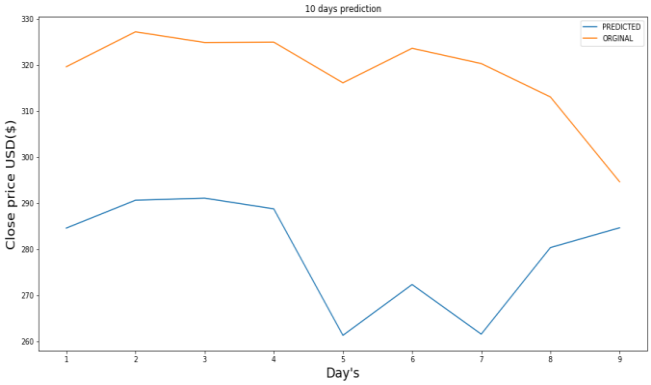

Fig. 8. Visualizing the APPLE 10 days prediction

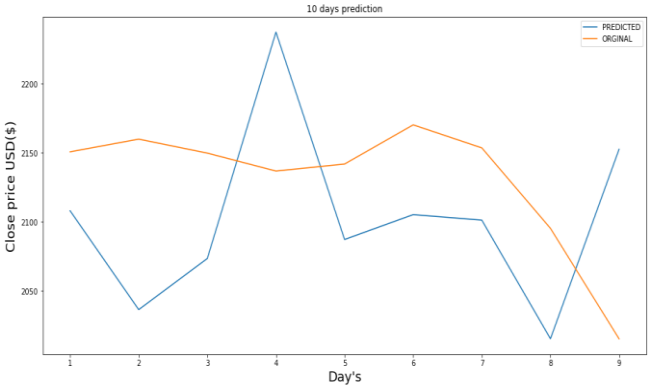

Fig. 9. Visualizing the AMZN 10 days prediction.

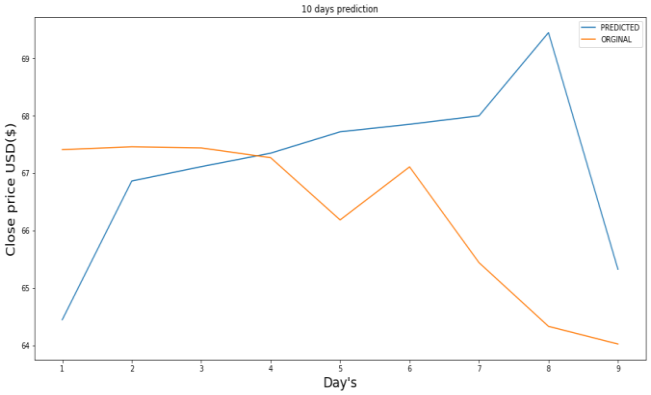

Fig. 10. Visualizing the INTC 10 days prediction.

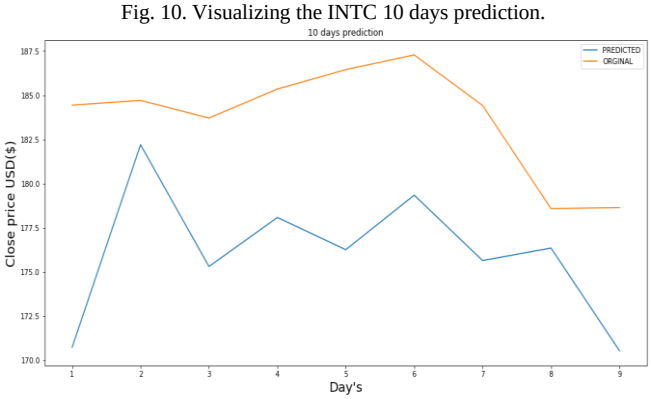

Fig. 11. Visualizing the MSFT 10 days Prediction.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

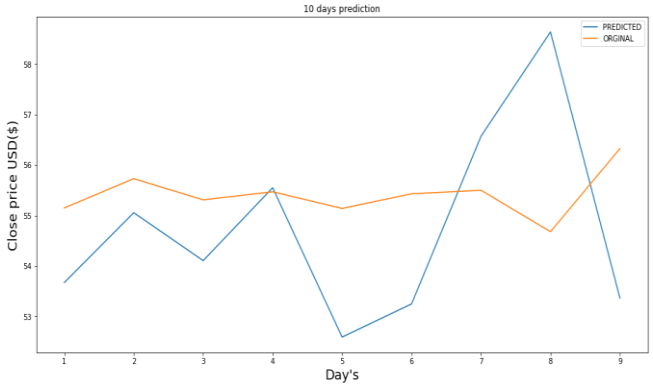

Fig. 12. Visualizing the ORCL 10 days Prediction.

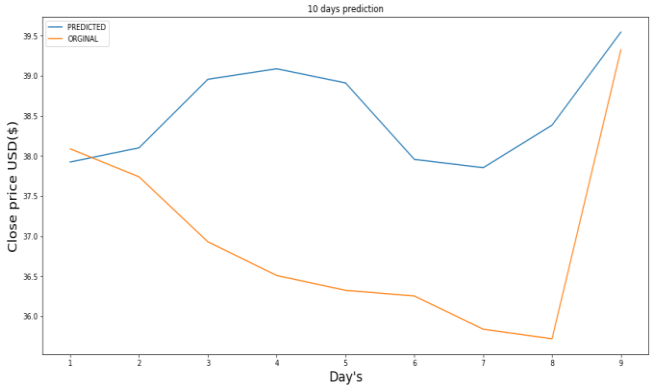

Fig. 13. Visualizing the PFE 10 days Prediction.

There we get the closest value for the model we have trained. For some companies, it predicts the close value and for some, it goes high value and low for some more but it overcomes the draw backs of the normal algorithm.

V. CONCLUSION

Stock price prediction is one of the not topics in that field of machine learning. The prediction of stock price not only helps developers also help the investors to invest in a profitable company and gain some profit. By using LSTM we get more accuracy than other algorithms in machine learning. Here we only considered the closing price of each day and created the model and get the closest predicted value. We also take 60 days of data for particular y days that also helped the model to recognize the pattern in the sequence data and predict the next one. We applied the same model on 7 top companies for 10 days that lead to getting a conclusion that for some company it gets the closest value and for some, it gets very large difference there we get 60% of closest prediction.

The further enhancement we try to add new feathers to the existing one that feather is news and sentiments of the country and company. These feathers increase the model to find the accurate value at add times.

REFERENCES

[1] Graves, A.; Liwicki, M.; Fernandez, S.; Bertolami, R.; Bunke, H.; *Schmidhuber, J.* (2009). *"A Novel Connectionist System for Improved Unconstrained Handwriting Recognition"*.
[2] Future Stock Price Prediction using LSTM Machine Learning Algorithm Mrs. Nivethitha1, Pavithra.V2, Poorneshwari. G3, Raharitha. R4.
[3] Stock Price Prediction Using LSTM on Indian Share Market ,Achyut Ghosh1, Soumik Bose1, Giridhar Maji2, Narayan C. Debnath3, Soumya Sen1.
[4] Stock Price Prediction Using Long Short Term Memory ,Raghav Nandakumar1, Uttamraj K R2, Vishal R3, Y V Lokeswari4.
[5] Illustrated Guide to LSTM's and GRU's: A step by step explanation Available:https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21
[6] Illustrated Guide to LSTM's and GRU's: A step by step explanation Available:https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21
[7] Ian Goodfellow, Yoshua Bengio, Aaron Courville - Deep Learning (2016) [Online]. Available:https://github.com/janishar/mit-deep-learning-book-pdf/blob/master/complete-book-pdf/Ian%20Goodfellow%2C%20Yoshua%20Bengio%2C%20Aaron%20Courville%20-%20Deep%20Learning%20(2017%2C%20MIT).pdf
[8] Stock market analysis using machine learning irjet 22 july 2019