IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11339

Stock Market Prediction using Machine Learning Algorithm

Akankshya Rout[1], Ayush Kumar Bar[2], Satya Priya Saha[3], Dr. Avijit Kumar Chaudhuri[4]

[1,2,3] UG – Computer Science and Engineering, Techno Engineering college Banipur, Kolkata, West Bengal

[4] Professor, Computer Science and Engineering, Techno Engineering college Banipur, Kolkata, West Bengal

Abstract: Stock market price data is huge and it changes every second. As it is a complex system in which people either make money or lose all their savings, hence it is important to understand the stock market. In the era of big and dynamic data, machine learning for predicting stock market prices and trends has become even more popular than ever. In this paper, we tried to predict the trend of the stock market. A model with a supervised machine learning algorithm is used to predict prices. We collected data of every company from the beginning from Yahoo finance and proposed comprehensive customization of RNN Machine Learning based models which are known as LSTM for predicting price trends of stock markets.

The proposed solution is comprehensive as it includes pre-processing of the stock market dataset, utilization of multiple feature engineering techniques, combined with an RNN based system for stock market price trend prediction.

In the yearly forecasting model, historical prices have been trained and achieved an accuracy of 84.0%.

We conducted comprehensive evaluations on frequently used machine learning models and concluded that our proposed solution outperforms due to the comprehensive feature engineering that we built.

Through our detailed design and evaluated prediction term lengths, feature engineering and data pre-processing methods, this work will help investors to invest in the stock by comparing stocks of different enterprises periodically, hence resulting in less risk. Also, it will contribute to the financial and technical domains of the stock analysis research community.

Keywords: Stock Market, Machine Learning, LSTM, RNN, Forecast, Feature Engineering

1. INTRODUCTION

Stock market prediction and analysis are one of the difficult tasks to do. There are various reasons for this, including the unsteady nature of the market and various dependent and independent variables that impact the worth of a specific stock in the market.

Nowadays various types of Machine Learning Algorithms have started taking place to analyse stock market data.

In synopsis, Machine Learning Algorithms are broadly used by numerous associations in Stock market prediction. Therefore, this paper will also help in understanding the stock market and growth of the different enterprises

1.1. Indian Stock Market Overview

According to the World Bank, the Indian economy is the third-largest in terms of purchasing power parity, which can still grow within the predictable future. That said, the country's booming economy is probably going to experience many ups and downs, as well as movements in its stock exchange, which can considerably impact its growth.

So, let's perceive how the stock exchange affects India' economy.

The markets get their unsteady nature from the value fluctuations of individual stocks. As prices increase or decrease, market volatility influences businesses and shoppers. Throughout a bull phase, the stock prices go up. Additionally, it helps the economy grow positively. Likewise, consumer defrayment conjointly rises as people become more optimistic concerning the market and purchase more products and services. So, businesses supplying these products and services begin to supply and sell more.

In every country there is at least one stock exchange market, where the stocks/shares of listed corporate sectors or enterprises purchase or sell. When an enterprise initially registers itself in any stock exchange to become a public company, the promoter groups sell a substantial number of shares to the public as per Government standards. When the promoter organization unloads the considerable number of shares to public retail investors, then at that point, those could be exchanged in the secondary market, i.e., stock exchange.

The two most efficacious stock exchanges are BSE (Bombay Stock Exchange) and NSE (National Stock Exchange) in India. According to [1] in the financial year 2020, a total of over 7,400 companies were listed in the NSE and BSE across India. Both the exchanges have comparative trading and market opening and shutting time which helps individual investors to participate in the share market conveniently.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11339

According to [2] India's benchmark 10-year bond yield rose to 6.73%, up 7 basis points from its previous close and its highest level since Dec 13, 2019. Until now in January, foreign investors have dumped $2.2 billion of Indian shares after having bought a net $3.76 billion in 2021. They had bought $23.29 billion worth shares in 2020 and $14.23 billion in 2019. They are still net buyers of $575.35 million worth debt so far this month after having sold $3.66 billion in 2021, which clearly shows that investing in Indian stocks is greatly profitable nowadays.

Stock market prediction and analysis are one of the difficult tasks to do. There are various reasons for this, including the unsteady nature of the market and various dependent and independent variables that impact the worth of a specific stock in the market.

Nowadays various types of Machine Learning Algorithms have started taking place to analyse stock market data.

In synopsis, Machine Learning Algorithms are broadly used by numerous associations in Stock market prediction. Therefore, this paper will also help in understanding the stock market and growth of the different enterprises

2. LITERATURE REVIEW

The prime objective of our research was to predict the future stock prices of Indian companies which helps in the growth of the stock market. As we studied the trends further, we came across a machine learning approach to predict and forecast future stock prices to help Indian investors.

After going through [3][6][7] research papers, articles and journals our key findings included the use of RNN's LSTM to overcome the challenges faced during model training. In many journals such as [2] which they had discovered correlation between" public sentiment" and" market sentiment" with an accuracy of 75.56%.

To train long term data, LSTM was a really good option because LSTM requires only about half of DNN and more parameters than CNN. LSTM's nature of being the slowest to train comes with its advantage for being able to take a look at longer sequences of inputs without expanding the network size.

3. LSTM ARCHITECTURE

Long Short-Term Memory Network is a high level RNN, a sequential network, that permits data to persevere. It is equipped for taking care of the disappearing gradient issue looked at by RNN. An intermittent neural network, otherwise called RNN is utilized for steady memory.

For example, suppose while watching a video you recall the past scene or while reading a book you realize what occurred in the prior part. Correspondingly RNNs work, they recollect the past data and use it for handling the current info. The shortcoming of RNN is, they cannot recall long-term conditions because of disappearing gradients. LSTMs are expressly intended to keep away from long-term reliance issues.

At an advanced level LSTM works a lot like a RNN cell. Here is the inside functioning of the LSTM network. Similar to RNN, an LSTM also contains a hidden state or short-term memory in which H(t-1) represents the hidden state of the previous timestamp and Ht is the hidden state of the current timestamp. Furthermore, LSTM also contains a cell state represented by C(t-1) for previous and C(t) for current timestamps respectively.

The LSTM comprises three sections, as shown in the picture below and each part performs an individual function.

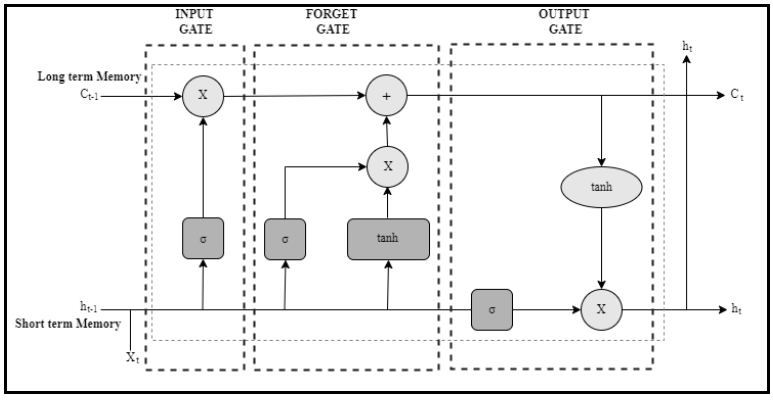

Fig. 1. LSTM Gate

- Forget Gate: - It picks whether the data coming from the previous timestamp is to be recollected or is insignificant and can be neglected.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11339

Equation for Forget Gate: -

$f_t = \sigma(x_t * U_f + H_{t-1} * W_f)$

Here,

- x_t =Input at the current timestamp
- U_f =weight associated with the input
- H_{t-1}=The hidden state at the previous timestamp
- W_f=It is the weight matrix associated with the hidden state

After multiplying ft with sigmoid function, it will make f_t a number somewhere in the range of 0 and 1. This ft is subsequently multiplied with the cell state of the previous timestamp as shown below.

$C_{t-1} * f_t = 0$ ...if $f_t = 0$ omit everything
$C_{t-1} * f_t = 0$ ...if $f_t = 1$ omit nothing

- Input Gate: - Here, the cell tries to read new information from the input to this cell.

It is used to evaluate the significance of the new information carried by the input.
Equation for Input gate: -
$i_t = \sigma(x_t * U_i + H_{t-1} * W_i$

Here,

- x_t = Input at the current timestamp t
- U_i = weight matrix of input
- H_{t-1} = A hidden state at the previous timestamp
- W_i = Weight matrix of input associated with hidden state

New information: -
Now the new information that should have been passed to the cell state is a function of a hidden state at the previous timestamp $t-1$ and input x at timestamp t. Here, the activation function is tanh. Because of the tanh function, the worth of new information will be between - 1 and 1. If N_t is negative the information will get subtracted from the cell state else if it is positive the information will get added to the cell state at the current timestamp.

$N_t = \tanh(x_t + U_c + H_{t-1} * W_c)(\text{new information})$

Nonetheless, the N_t won't be added straight to the cell state. Hence, the equation is

$C_t = f_t * C_{t-1} + i_t * N_t$ (updating cell state)

Where C_{t-1} is the cell state of the current timestamp.

Output Gate: - the cell transmits the updated information from the current timestamp to the next timestamp.

Equation for Output gate: -

$o_t = (x_t + U_o + H_{t-1} * W_o)$

Its value will also lie somewhere in the range of 0 and 1 because of this sigmoid function. Presently to figure out the current state we will utilize o_t and the tanh of the updated cell state. As shown below.
$H_t = o_t * \tanh(C_t)$

After summarizing it turned out, the hidden state is a function of long-term memory (C_t) and the current output. To get the output of the current timestamp implement the Adam activation on the hidden state H_t.

Output=adam(H_t)

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11339

4. METHODOLOGY

The motivation behind our framework is to predict the future share values of various enterprises and compute the future growth of the enterprises in different periods of time. Then, at that point, we dissect the prediction error for each enterprise in various sectors. Based on that we can easily compare the future stock prices of different companies for a short period of time.

We initially anticipate the future shutting price of 4 unique enterprises from a few pre-chosen areas with the help of LSTM. Two unique models have been built to predict stock market trends. First model anticipates the stock market trend for

the upcoming day (Daily prediction model) by considering all available data on a daily basis as input. Second model predicts

the stock market trend for the upcoming week/month by considering available data on yearly or monthly basis.

One of the statistical arguments considered is the relationship between trend of a day and closing price of stock traded on the same day.

The forecast will be done on historical data and the future anticipation will be done for 1 month, 6 months & 1 years. In these three different time spans (1 month, 6 months & 1 year), we calculate the growth of those companies. Then by analysing the variation of shutting price for each time period, we compare the companies which have maximum growth, i.e., less error for the particular sector.

4.1. Proposed Method

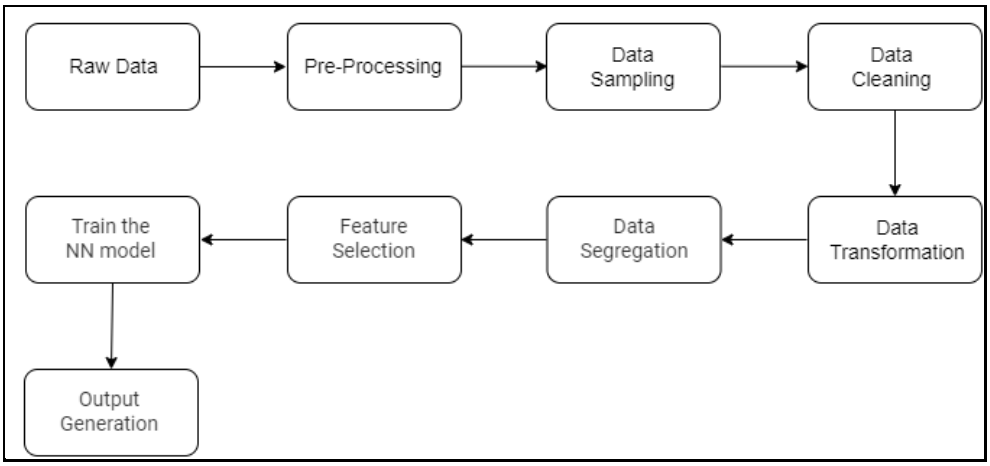

Fig. 2. Block Diagram

4.2. Implementation Steps

Step1: Raw Stock price dataset:

Day-wise past stock prices of selected enterprises are gathered from the BSEINDIA official site. Then it is cleaned using data cleaning methods and removing the null and empty spaces.

Step2: Pre-processing:

This includes the following steps:

Data sampling: Reducing a part of data to use specific data specially for numerical data

Data cleaning: Removing the missing and null values.

Data Transformation: Normalizing the data to reduce bias with bigger numbers.

Data Segregation: Segregation of data into train and test set for evaluation

Step3: Feature Selection:

Here we are selecting the features (i.e., Date and Close) that are to be fed into the neural network.

Step4: Train the NN model:

The Neural Network is trained using the pre-processed training dataset. Proposed LSTM model consists of a sequential input layer followed by 1 LSTM layer and then a dense layer with activation.

Step5: Output Generation

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11339

4.3. Pseudocode

```
model = Sequential ()
model.add (LSTM (units=50, return_sequences=True,input_shape=(x_train.shape[1],1)))
model.add (LSTM (units=50, return_sequences=False))
model.add (Dense (units=25))
model.add (Dense (units=1))
model.compile (optimizer='adam', loss='mean_squared_error')
model.fit (x_train, y_train,validation_data=(x_test, y_test), epochs=20,verbose=0)
```

5. RESULTS AND DISCUSSION

The recommended LSTM based model is carried out with Python. In the given Table 1 the Accuracy, Misclassification, Precision, Sensitivity, Specificity values for
various organizations belong to the IT Sector based on the recorded information of the 1 Year is shown.

Table.1.

	Accuracy	Misclassification	Precision	Sensitivity	Specificity
Amazon	0.84	0.152	0.55	0.55	0.91
Google	0.601	0.397	0.32	0.32	0.72
Microsoft	0.605	0.394	0.32	0.32	0.72
Apple	0.685	0.315	0.37	0.37	0.79

As it very well may be seen, the Accuracy for each enterprise is between 60-80% which shows that the LSTM model is good at distinguishing relationships and patterns between variables in a dataset input and training. Consequently, the model is capable of generalizing the 'unseen' data, and thus better predictions and insights can be produced.

Table.2.

	1 month	6 month	1 year
Amazon	0.64	0.64	0.84
Google	0.63	0.63	0.601
Microsoft	0.64	0.63	0.605
Apple	0.65	0.64	0.685

From the above table 2 Accuracy table is given for 1 month, 6 month and 1 year for the 4 IT sectors.

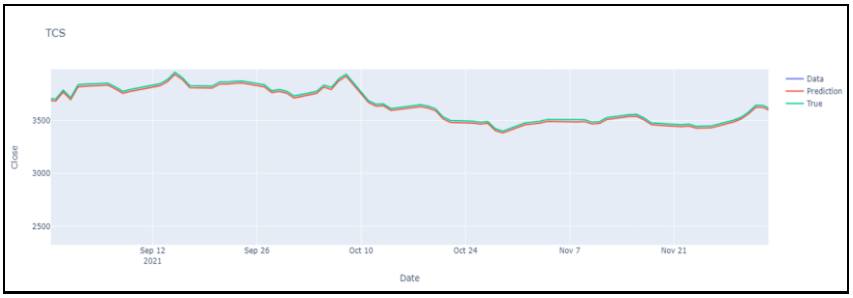

Fig.3. Prediction on test data

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11339

6. CONCLUSION AND FUTURE SCOPE

As observed in the past few years, investing has become very popular in recent years. Investing in the stock market is beneficial with high returns in the short term as well as long term period. But it also carries the risk of losing investment. We have many prediction models available in the market which can predict the stock price trend on a weekly basis. So, we tried to propose a similar kind of model with some different parameters with a good amount of accuracy.
This model can help in getting a better idea of highs and lows during purchasing or selling of stocks.
We can use this model for future studies and prediction of prices of crypto-currencies.

7. REFERENCES

[1] https://economictimes.indiatimes.com/
[2] https://www.irjet.net/archives/V5/i3/IRJET-V5I3788.pdf
[3]https://www.researchgate.net/publication/321503983_Stock_price_prediction_using_LSTM_RNN_and_CNN-sliding_window_model
[4] https://iopscience.iop.org/article/10.1088/1757-899X/790/1/012109/pdf
[5] https://www.ijcrt.org/papers/IJCRT2102617.pdf
[6]https://www.researchgate.net/publication/327967988_Predicting_Stock_Prices_Using_LSTM
[7]https://www.researchgate.net/publication/348390803_Stock_Price_Prediction_Using_LSTM

