



# STOCK MARKET PREDICTION USING MACHINE LEARNING

MR.KINGSLY STEPHEN<sup>1</sup>,O.BHARGAV SATHWIK<sup>2</sup>,HEMANTH.T<sup>3</sup>, G.DURGA  
DHIKSHITH<sup>4</sup>.

<sup>1</sup>Assistant Professor, Department of Computer Science and Engineering ,SRMIST Ramapuram Chennai.

<sup>2</sup>Department of Computer Science and Engineering, SRMIST Ramapuram,Chennai.[UG Scholar]

<sup>3</sup>Department of Computer Science and Engineering, SRMIST Ramapuram,Chennai.[UG Scholar]

<sup>4</sup>Department of Computer Science and Engineering, SRMIST Ramapuram,Chennai.[UG Scholar]

## 1.ABSTRACT: -

The act of attempting to anticipate the future value of a business stock or other financial instrument traded on an exchange is known as stock market prediction. A successful forecast of a stock's future price could result in a large profit. The factors that generally considers to predict the stock price are open, close, low, high, volume etc. The main objective of this project is to build a Machine learning model that will help to predict the future value of a company's stock. For Building this model we can use different Machine learning algorithms like SVM, naive Bayes, random forest and linear regression etc. Here in this project, as the stock market is volatile in nature, it will be effective if we build the model using LSTM. This LSTM is able to store the past information, which is very important in our project, because the previous price of stock is crucial in predicting its future price. Here the LSTM is an artificial recurrent neural network used in the field of Deep learning. For building this model we are using a well labelled dataset.

Keywords: - Machine learning, Deep learning, SVM, Naive Bayes, Random Forest, Linear Regression, LSTM.

## 2.INTRODUCTION

The stock market is a marketplace for investors to buy and sell investments, most often stocks, which are ownership shares in a publicly traded firm. An initial public offering, or IPO, is when a company sells shares of its stock on a stock exchange. Investors buy those shares, allowing the company to raise funds to expand its operations. The exchange then tracks the supply and demand of each listed stock, allowing investors to purchase and sell these stocks among themselves.

If someone in India wishes to list their company on the stock exchange, they have three possibilities. The Bombay Stock Exchange (BSE), the National Stock Exchange (NSE), and the Calcutta Stock Exchange are the three stock exchanges in India (CSE). The Sensex and the Nifty Fifty are two popular indexes that measure the market's overall trend.

Stock market forecasting and analysis are two of the most difficult things to complete. There are various reasons for this, including market volatility and a slew of other dependent and independent elements that go into determining the market value of a particular stock. These characteristics make it extremely difficult for any stock market analyst to accurately predict the rise and fall of the market.

Every day, millions of people all over the world invest in the stock market. Investors, management, and decision-makers will benefit from a solid stock price prediction model since it will assist them in making accurate and effective judgments

For Building this prediction model, we are using LSTM method because it is able to store the past information which is crucial in predicting the future price of stock.

For this project we are going to collect the dataset which contains the features like open price, low price, high price, closed price, number of trends etc. The scale of the dataset for the stock market we collected is fairly large and random, so we need models that are efficient and can handle the complexity of such a large amount of data. Due to hidden patterns, the stock data is convoluted and difficult to comprehend, so we should use machine learning techniques as they have ability to deal with such complex data. When we are going to build the model the open price, low price, high price, number of trends are considered as input independent variables and closed price is considered as target variable for our project.

### **3.RELATED WORK: -**

In the work of Shekar Singh and Seema Sharma[2], the researchers have used multivariate regression technique in the form of partial least squares regression (PLSR). Researchers used the PLSR method to make predictions in this study, which proved to be highly effective. One thing that can be said about the many models that exist is that no single model utilising a single algorithm can accurately anticipate trends for every market, some models may successfully predict some stocks, while others may correctly predict others. so the major drawback for this work is that they have used single algorithm for building the model.

In the work of Upendra Singh and Ashish Sharma[1], the researchers have used different regression methods like linear, polynomial and RBF to build the predictive model for stock market prediction. Results of these methods are compared on the basis of confidence value, and they came to conclusion that linear regression is the best predictive model. But the drawback with this approach is that this model can't store the past information which is very important to predict the future value of a company's stock price.

## 4. METHODOLOGY

### 4.1 Data collection and preparation: -

For doing this project first we collect the historical stock data from companies of different sectors. This collected dataset contains the features like low price, high price, close price, open price, volume etc.

In this project we take low, high, and open price are taken as input independent variables and closed price is chosen as target dependent variable. In this particular project first, we will train the model using the dataset and then we will test the model.

Here the collected dataset is divided into two parts that is training set and testing set. The training set is used to train the model and the testing set is used to test the model. We have divided the data in 80:20 ratio i.e. 80 percentage of data is used for training and 20 percentage of data is used for testing.

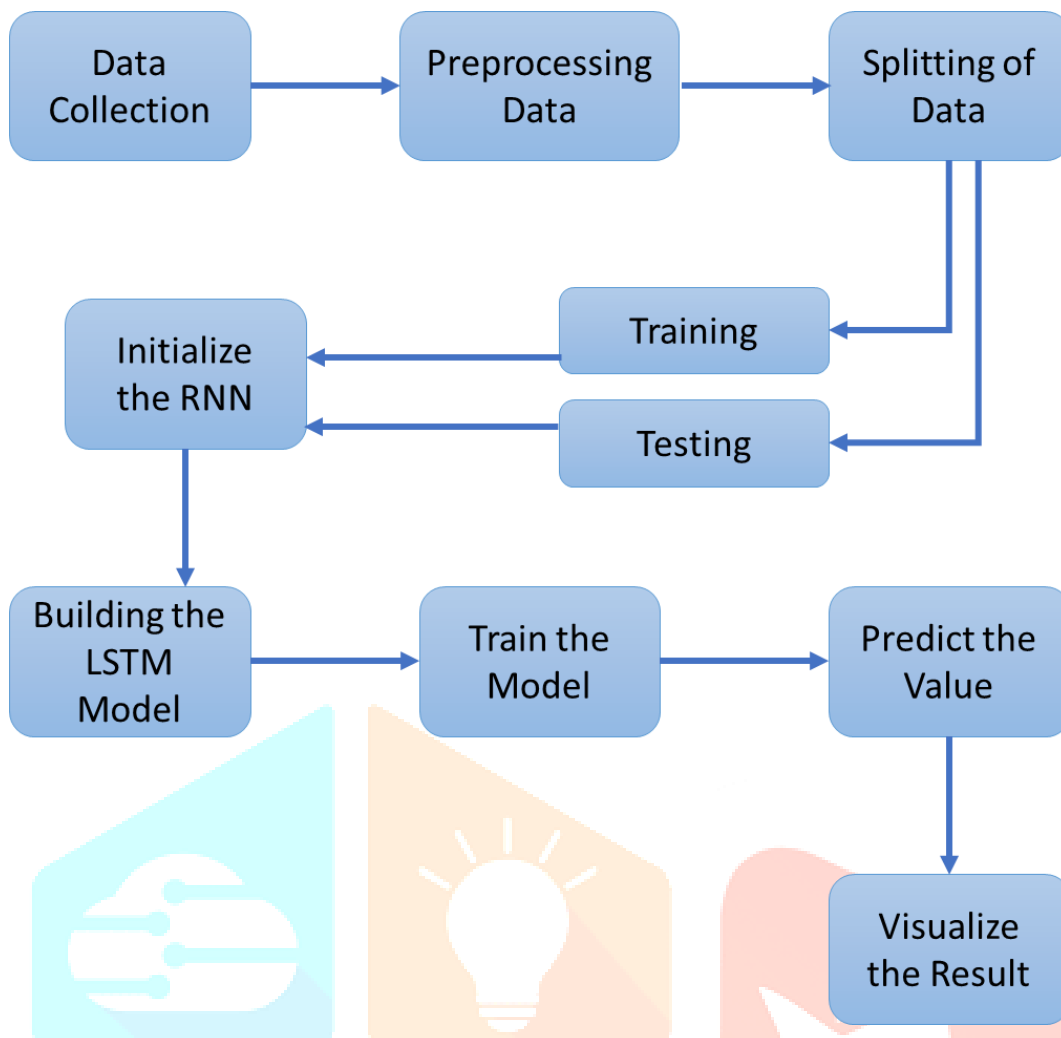
Before splitting the data for training and testing, we should pre-process the collected data because, the collected data may contain some outliers, missing values and noisy values. To overcome these issues and to clean the data we will do pre-processing.

And then we will build the model and then do training and testing the model etc.

## 5. PROPOSED SYSTEM

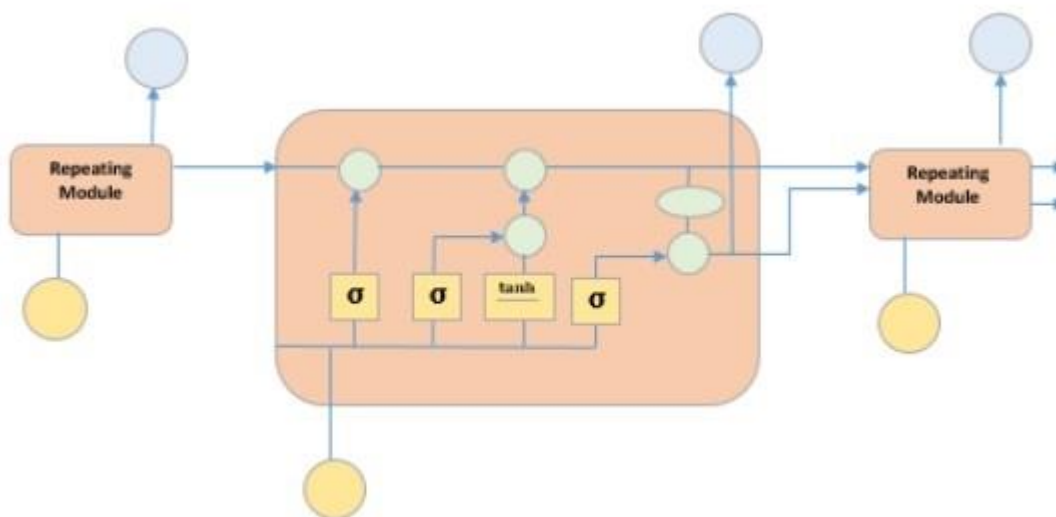
### 5.1 workflow: -

In this section we will know about the work flow of our project. First, we will collect the data then we will pre-process the data and then we will divide the data into training set and testing set. And after building the model we will train and then do the testing and finally we visualise the data. This is the work flow of the project and we can get the clear view of the workflow from below diagram



### 5.2 Long short term Memory (LSTM):-

It's a unique type of recurrent neural network that can learn long-term data relationships. This is possible because the model's recurring module is made up of four layers that interact with one another.



Four neural network layers are depicted in yellow boxes, point wise operators in green circles, input in yellow circles, and cell state in blue circles in the diagram above. An LSTM module has a cell state and three gates, giving it the ability to learn, unlearn, or retain information from each of the units selectively. By allowing only a few linear interactions, the cell state in LSTM allows information to travel across the units without being altered. Each unit contains an input, output, and a forget gate that adds or removes data from the cell state. The forget gate utilises a sigmoid function to determine whether information from the previous cell state should be ignored. The input gate uses a point-wise multiplication operation of 'sigmoid' and 'tanh' to control the information flow to the current cell state. Finally, the output gate determines which data should be sent to the next concealed state.

### 5.3 LSTM Model Construction: -

We'll need to import the Keras library and packages to create the LSTM model. Keras is a high-level API for constructing and training models based on tensor flow. We'll be importing various libraries here.

Sequential - It's a linear stack of layers that may be used to construct a sequential layer by running a list across it.

Dense - This is the standard layer of a densely connected neural network. It is the most widely utilised layer for changing the output dimension. It is a metrics vector multiplication that updates the value of the trainable parameter during back propagation.

Following that, we'll set up RNN. We'll use a regression model to solve time series problems like this. The initial stage in this process is to read the data, which is sequential, and then assign it to the regressor. A sequential input layer precedes three LSTM layers and a dense layer in this LSTM model.

After building of LSTM model, the important step is to train the model by using the collected dataset and then do the testing.

After the testing we use Root mean square error (RMSE) method to determine how accurate our model is predicting. The low RMSE value indicates the good prediction of model.

## 6.CONCLUSION

In this paper we have discussed how the stock price prediction model is build based on the LSTM Approach. This model works very well and gives good prediction on future value of a company's stock. This approach works well, because it has the ability to store past information which is very useful for predicting the future price of a share. We have utilised the Machine learning approach and used LSTM (Deep learning Model) to build our model. The tool we used here is jupyter notebook and also, we have derived several libraries like scikit learn, keras etc that are useful for our project. And also, RMSE is used to find how good our model is predicting.

## REFERENCES

- [1]. Ashish Sharma & Upendra Singh (2017) International Conference on Electronics, Communication and Aerospace Technology  
ICECA 2017
- [2]. Shekar Singh & Seema Sharma, Forecasting Stock Price Using Partial Least Squares regression, IEEE (2018)
- [3] H. Wang, G. Wang, G. Li, J. Peng, and Y. Liu, "Deep belief network based deterministic and probabilistic wind speed forecasting approach," Applied Energy, vol. 182, pp. 80–93, 2016.
- [4] D. K. Kılıc, and O. Uğur, "Multiresolution analysis of S&P500 time series," Annals of Operations Research, pp. 1–20, 2016.
- [5] P. Li, C. Jing, T. Liang, M. Liu, Z. Chen, and L. Guo, "Autoregressive moving average modeling in the financial sector," in 2nd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), 2015, pp. 68–71.
- [6] G. Zhang, X. Zhang, and H. Feng, "Forecasting financial time series using a methodology based on autoregressive integrated moving average and Taylor expansion," Expert Systems, vol. 33, no. 5, pp. 501–516, 2016.
- [7] M. Bildirici, O. Ersin et al., "Nonlinearity, volatility and fractional integration in daily oil prices: smooth transition autoregressive-st-fi (ap) garch models," Romanian Journal of Economic Forecasting, vol. 3, pp. 108–135, 2014

