IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# STOCK MARKET PREDICTION USING MACHINE LEARNING

[1]Sathya S, [2] Dr. N. Radha

[1]Student,
[2]Associate Professor,
[1]Department of Data Analytics,
[1]PSGR Krishnammal College for Women, Coimbatore, Tamilnadu, India.

**ABSTRACT**

Forecasting of stock prices has been a challenging task for many researchers and analysts. The profits can be maximized if the market value is predicted keeping the risk low. Nowadays, investors are highly immersed in the research area of stock price prediction. To make a good and successful investment, many investors are eager in knowing the future of the stock market. Appealing and effective prediction systems help traders, investors, and analyst by providing supportive information like the future direction of the stock market. Long Short-Term Memory (LSTM) approach is widely used to predict stock market indices. Accurate and adequate stock market prediction is of great interest to investors; However, stock markets are driven by volatile factors such as blogs and news that make it hard to predict stock market index based on the historical data. Stock market price can be predicted using machine learning algorithms on information contained on social media and financial news, as this data can change investors' behavior in the decision making related to stock markets. LSTM is better for time series forecasting models that can predict future values based on previous, sequential data. So, this model provides greater accuracy for demand forecasters which helps in making better and best decisions.

***Index Terms* - LSTM, time-series, forecasting.**

## I. INTRODUCTION

The stock market broadly refers to the collection of exchanges and other venues where the buying, selling, and issuance of shares of publicly held companies take place. Such financial activities are conducted through institutionalized formal exchanges (whether physical or electronic) or via over-the-counter (OTC) marketplaces that operate under a defined set of regulations. While both the terms "stock market" and "stock exchange" are often used interchangeably, the latter term generally comprises a subset of the former. If one trades in the stock market, it means that they buy or sell shares on one (or more) of the stock exchange(s) that are part of the overall stock market.

LSTMs are very powerful in sequence prediction problems because they're able to store past information. This is important in our case because the previous price of a stock is crucial in predicting its future price. The main objective of this internship is to forecast the stock prices using LSTM. While predicting the actual price of a stock is an uphill climb, we can build a model that will predict whether the price will go up or down. Also, some authors overviewed the forecasting of returns using of ADaBoost algorithm [11]. Others proceeds to forecast stock returns using unique decision-making model for day trading investments on the stock market the model developed by the authors use the support vector machine (SVM) method, and the mean- variance (MV) method for portfolio selection [8]. Another paper conversed deep learning models for smart indexing [5]. Also, some study has covered a large number of trends and Applications of Machine Learning in Quantitative Finance [2], the literature review covered by this paper consist of return forecasting portfolio construction, ethics, fraud detection, decision making, language processing and sentiment analysis. These models don't depend one long term memory (passed sequences of data), in this regard a class of machine learning algorithms based on Recurrent Neural Network prove to be very useful in financial market price prediction and forecasting.

## II. LITERATURE REVIEW

There are lots of research work in stock market prediction as well as in LSTM. Almost every data mining and prediction techniques were applied for prediction of stock prices. Many different features and attributes were used for the same purpose. There are three main categories of stock market analysis and prediction such as (a) Fundamental analysis, (b) Technical analysis and (c) Time series analysis.

Gers & Schmidhuber [2] proposed a model on variation of LSTM by introducing peephole connections. In this model, the gate layers can look into the cell state. In this case, decision to add new information or to forget it is taken together. It forgets only when it needs to input something in its place. This architecture inputs new values to the cell state when it forgets anything older.

Hiransha [4] proposed a model, employed three different deep learning network achitectures such as RNN, CNN and LSTM to forecast stock price using day wise past closing prices. They have considered two company from IT sector (TCS and Infosys) and one from Pharma sector(Cipla) for experiment. The uniqueness of the study is that they have trained the models using data from a single company and used those models to predict future prices of five different stocks from NSE and NYSE (Newyork Stock Exchange). They argued that linear models try to fit the data to the model but in deep networks underlying dynamic of the stock prices are unearthed. As per their results CNN outperformed all other models as well as classical linear models. The DNN could forecast NYSE listed companies even though the model has learned from NSE dataset. The reason could be the similar inner dynamics of both the stock exchanges.

Kim [5] proposed a model on, Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data. The feature fusion long short-term memory-convolutional neural network (LSTM-CNN) model, that combines features learned from different representations of the same data, namely, stock time series and stock chart images, to predict stock prices. The model is used to utilize for extracting temporal features and image features. The performance of the proposed model relative to those of single models (CNN and LSTM) using SPDR S&P 500 ETF data is measured. The feature fusion LSTM-CNN model outperforms the single models in predicting stock prices. The candlestick chart is the most appropriate stock chart image to use to forecast stock prices. The minute-by-minute SPY ticker data, which has the largest trading volume among ETF markets was used here.

Shengyi Zhao [8] proposed a model on, Tomato Leaf Disease Diagnosis Based on Improved Convolution Neural Network by Attention Module. A deep convolutional neural network that integrates an attention mechanism, which can better adapt to the diagnosis of a variety of tomato leaf diseases is being proposed. This model can accurately extract complex features of various diseases which has average identification accuracy of 96.81% on the tomato leaf diseases dataset. model provides a high-performance solution for crop diagnosis under the real agricultural environment. The image data of tomato leaf health and disease in this paper comes from the Plant Village open source database. The database contains a large number of plant disease images and is the world's largest crop database.

## III. METHODOLOGY

To The data collection is done with the range from 2010 till 2021 which is taken from yahoo finance. Using a Keras Long Short-Term Memory (LSTM) Model to Predict Stock Prices. LSTMs are very powerful in sequence prediction problems because they're able to store past information. This is important in our case because the previous price of a stock is crucial in predicting its future price.

**DATASET DESCRIPTION**

**DATE:** This attribute contains the date of the stocks.
**OPEN:** It states the value of opening stock price for each day.
**HIGH:** This is the highest stock price of the day.
**LOW:** This attribute states the lowest stock price of the day.
**CLOSE:** This represents the value of closing stock price for each day.
**ADJ CLOSE:** This attribute states the adjacent value of closing stock price of the day.
**VOLUME:** It represents the volume of stocks being traded for the day.

| Date | Open | High | Low | Close | Adj Clos | Volume |
|---|---|---|---|---|---|---|
| 5/17/2021 | 3055 | 3075.2 | 3036 | 3069.75 | 3036.981 | 2890462 |
| 5/18/2021 | 3100 | 3124 | 3078 | 3088.8 | 3055.828 | 2098538 |
| 5/19/2021 | 3084 | 3118 | 3067.1 | 3082 | 3049.1 | 1986041 |
| 5/20/2021 | 3067.1 | 3088.8 | 3052.1 | 3060 | 3027.335 | 2329027 |
| 5/21/2021 | 3061 | 3088.2 | 3055.1 | 3080.5 | 3047.616 | 1685566 |
| 5/24/2021 | 3081.5 | 3105 | 3072 | 3081.5 | 3048.606 | 1652260 |
| 5/25/2021 | 3092 | 3128.25 | 3082.1 | 3114 | 3095.828 | 1841613 |
| 5/26/2021 | 3120 | 3165 | 3103.8 | 3158.5 | 3140.069 | 1923753 |
| 5/27/2021 | 3161.95 | 3217.75 | 3161.8 | 3180 | 3161.443 | 5959785 |
| 5/28/2021 | 3189.5 | 3198 | 3135.65 | 3143.6 | 3125.256 | 1763701 |
| 5/31/2021 | 3150 | 3170.35 | 3128.6 | 3159.15 | 3140.715 | 1652799 |
| 6/1/2021 | 3168.6 | 3169.95 | 3132 | 3153 | 3134.601 | 1377441 |
| 6/2/2021 | 3150.85 | 3159.45 | 3115 | 3129.45 | 3111.188 | 2240078 |
| 6/3/2021 | 3154.55 | 3154.55 | 3122.65 | 3141.25 | 3122.919 | 1281706 |
| 6/4/2021 | 3128 | 3156.85 | 3125 | 3143.75 | 3125.405 | 1836060 |
| 6/7/2021 | 3145 | 3190.45 | 3133.6 | 3183.2 | 3164.625 | 2559821 |
| 6/8/2021 | 3198 | 3231 | 3187.2 | 3200.15 | 3181.476 | 2574057 |
| 6/9/2021 | 3202 | 3220 | 3186.2 | 3200.25 | 3181.575 | 1710925 |
| 6/10/2021 | 3210 | 3224 | 3193 | 3216.8 | 3198.029 | 1874324 |
| 6/11/2021 | 3211.55 | 3309 | 3211 | 3273.8 | 3254.696 | 3452954 |
| 6/14/2021 | 3279 | 3308.7 | 3270.35 | 3276.35 | 3257.231 | 1847360 |
| 6/15/2021 | 3298 | 3298 | 3251.55 | 3262.75 | 3243.71 | 1304985 |
| 6/16/2021 | 3262.1 | 3294.7 | 3253 | 3274.35 | 3255.243 | 1635552 |
| 6/17/2021 | 3265.5 | 3336.05 | 3260 | 3317.75 | 3298.389 | 2273413 |
| 6/18/2021 | 3350.9 | 3358 | 3275 | 3297.3 | 3278.059 | 3380431 |
| 6/21/2021 | 3265 | 3286 | 3251.7 | 3273.1 | 3254 | 1130569 |
| 6/22/2021 | 3304 | 3327.05 | 3285 | 3301.2 | 3281.936 | 1708688 |
| 6/23/2021 | 3329 | 3329 | 3256.4 | 3261.4 | 3242.368 | 1467104 |

**Figure 3.1 : Sample dataset**

## IV. PROCESS FLOW

### BUILDING A MODEL

5 steps in the LSTM model life-cycle in Keras that we are going to look at.

1. Defining the Network
2. Compiling the Network
3. Fitting the Network
4. Evaluating the Network
5. Make Predictions

**Defining a Network**

The first step is to create an instance of the Sequential class. Then you can create your layers and add them in the order that they should be connected. The LSTM recurrent layer comprised of memory units is called LSTM(). A fully connected layer that often follows LSTM layers and is used for outputting a prediction is called Dense().

**Compiling a Network**

Compilation is an efficiency step. It transforms the simple sequence of layers that we defined into a highly efficient series of matrix transforms in a format intended to be executed on your GPU or CPU, depending on how Keras is configured. Think of compilation as a precompute step for your network. It is always required after defining a model. Compilation requires a number of parameters to be specified, specifically tailored to training your network. Specifically, the optimization algorithm to use to train the network and the loss function used to evaluate the network that is minimized by the optimization algorithm.

**Fitting into the Network**

Once the network is compiled, it can be fit, which means adapt the weights on a training dataset. Fitting the network requires the training data to be specified, both a matrix of input patterns, X, and an array of matching output patterns, y. The network is trained using the backpropagation algorithm and optimized according to the optimization algorithm and loss function specified when compiling the model. The backpropagation algorithm requires that the network be trained for a specified number of epochs or exposures to the training dataset. Each epoch can be partitioned into groups of input-output pattern pairs called batches. This defines the number of patterns that the network is exposed to before the weights are updated within an epoch. It is also an efficiency optimization, ensuring that not too many input patterns are loaded into memory at a time.

**Evaluating the Network**

We can evaluate the performance of the network on a separate dataset, unseen during testing. This will provide an estimate of the performance of the network at making predictions for unseen data in the future. The model evaluates the loss across all of the test patterns, as well as any other metrics specified when the model was compiled, like classification accuracy. A list of evaluation metrics is returned.

## V. RESULT AND ANALYSIS

**Make Predictions**

In the case of a regression problem, these predictions may be in the format of the problem directly, provided by a linear activation function. For a binary classification problem, the predictions may be an array of probabilities for the first class that can be converted to a 1 or 0 by rounding. For a multiclass classification problem, the results may be in the form of an array of probabilities (assuming a one hot encoded output variable) that may need to be converted to a single class output prediction using the argmax() NumPy function. Alternately, for classification problems, we can use the predict_classes() function that will automatically convert uncrisp predictions to crisp integer class values.

```
In [59]: #Get the quote
quote = web.DataReader('AAPL', data_source='yahoo', start='2010/08/26', end='2021/10/11')

#Create a new dataframe
new_df = quote.filter(['Close'])

#Get the last 60 day closing price values and convert the dataframe into an array
last_60_days = new_df[-60:].values

#Scale the data to be values between 0 and 1
last_60_days_scaled = scaler.transform(last_60_days)

#Create an empty list
X_test = []

#Append the past 60 days
X_test.append(last_60_days_scaled)

#Convert the X_test data set to a numpy array
X_test = np.array(X_test)
#Reshape the data
X_test = np.reshape(X_test, (X_test.shape[0], X_test.shape[1], 1))

#Get the predicted scaled price
pred_price = model.predict(X_test)

#Undo the scaling
pred_price = scaler.inverse_transform(pred_price)
print(pred_price)

[[134.00246]]
```

**Figure 4.1 : Predicted price on 11-10-2021**

| DATE | CLOSE | PREDICTIONS |
|---|---|---|
| 2019-07-23 | 52.209999 | 50.598888 |
| 2019-07-24 | 52.167500 | 50.830814 |
| 2019-07-25 | 51.755001 | 51.060478 |
| 2019-07-26 | 51.935001 | 51.183311 |
| 2019-07-29 | 52.419998 | 51.270836 |
| … | … | … |
| 2021-10-04 | 139.139999 | 134.780914 |
| 2021-10-05 | 141.110001 | 133.940231 |
| 2021-10-06 | 142.000000 | 133.474579 |
| 2021-10-07 | 143.289993 | 133.379944 |
| 2021-10-08 | 142.899994 | 133.677750 |

**Table 4.1 : Predicted price on 11-10-2021**

## CONCLUSION

The popularity of stock market trading is growing rapidly, which is encouraging researchers to find out new methods for the prediction using new techniques. The forecasting technique is not only helping the researchers but it also helps investors and any person dealing with the stock market. In order to help predict the stock indices, a forecasting model with good accuracy is required. In this work, we have used one of the most precise forecasting technology using Recurrent Neural Network and Long Short-Term Memory unit which helps investors, analysts or any person interested in investing in the stock market by providing them a good knowledge of the future situation of the stock market.

## REFERENCES

[1] F. a. o. Eugene (1970), "Efficient capital markets: a review of theory and empirical work," Journal of finance, vol. 25( 2) : 383-417.
[2] F. A. Gers and J. Schmidhuber (2000), “Recurrent nets that time and count”, in Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, IEEE : 189-194.
[3] Heaton, J. B., Polson, N. G., & Witte, J. H. (2017), “Deep learning for finance: deep portfolios”. Applied Stochastic Models in Business and Industry, 33(1) : 3-12.
[4] Hiransha Ma , Gopalakrishnan E.Ab , Vijay Krishna Menonab, Soman K.P(2018), “NSE Stock Market Prediction Using Deep-Learning Models”, Procedia Computer Science 132 : 1351–1362.
[5] T. Kim and H. Y. Kim (2019), "Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data," PloS one, vol. 14(2) : e0212320.
[6] Paiva, F. D., Cardoso, R. T. N., Hanaoka, G. P., & Duarte, W. M. (2018). “Decision-Making for Financial Trading: A Fusion Approach of Machine Learning and Portfolio Selection”. Expert Systems with Applications.
[7] M. Roondiwala, H. Patel and S. Varma (2017), "Predicting stock prices using LSTM," International Journal of Science and Research (IJSR), vol. 6(4) : 1754-1756.
[8] Shengyi Zhao, Yun Peng, Jizhan Liu * and Shuo Wu (2011) , “Tomato Leaf Disease Diagnosis Based on Improved Convolution Neural Network by Attention Module”, Attention module, Agriculture.
[9] Md. Arif Istiake Sunny; Mirza Mohd Shahriar Maswood; Abdullah G. Alharbi (2020), “Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model”, 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES).
[10] Khaled A. Althelaya; El-Sayed M. El-Alfy; Salahadin Mohammed (2020), “Evaluation of bidirectional LSTM for short-and long-term stock market prediction”, 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES).
[11] Sreelekshmy Selvin; R Vinayakumar; E. A Gopalakrishnan; Vijay Krishna Menon; K. P. Soman (2017), “Stock price prediction using LSTM, RNN and CNN-sliding window model”, International Conference on Advances in Computing, Communications and Informatics (ICACCI).
[12] Kai Chen; Yi Zhou; Fangyan Dai (2015) , “A LSTM-based method for stock returns prediction: A case study of China stock market”, 2015 IEEE International Conference on Big Data (Big Data).
[13] David M. Q. Nelson; Adriano C. M. Pereira; Renato A. de Oliveira (2017), “Stock market's price movement prediction with LSTM neural networks”, 2017 International Joint Conference on Neural Networks (IJCNN).
[14] Chun Yuan Lai; Rung-Ching Chen; Rezzy Eko Caraka, “Prediction Stock Price Based on Different Index Factors Using LSTM” 2019 International Conference on Machine Learning and Cybernetics (ICMLC).
[15] AnitaYadav[a]C KJha[a]AditiSharan[b] (2020), “Optimizing LSTM for time series prediction in Indian stock market”, Procedia Computer Science,Volume 167: Pg 2091-2100.