Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 7, Issue 04, April -2020

A Survey Paper on Prediction of Stock priceUsing Machine Learning Technique

Maithili D. Patel[1], Prof. Prashant B. Swadas[2], Prof.Kirtikumar J. Sharma[3]

[1,2,3] *Department of Computer Engineering,Birla VishvakaramMahavidyalaya, Engineering College, VallabhVidyanagar, Gujarat, India*

Abstract - *The movement of the stock market is analysed and it is forecasted to gather the information and guidance for an investor when to buy or sell. It will also helpful for investors to gain more profit over their shares. This paper presents a general survey of Machine learning techniques. In the first paper some existing techniques to predict the target values are discussed, i.e.: LSTM, ARIMA, SVM. Further, these three methods are discussed in details. The mean of the target value is the price of shares the next day. After that, all the three existing methods are discussed and researched that best from three algorithms are LSTM, SVM, the best prediction model.*

Keywords-*Stock Market, Machine Learning, LSTM(Long Short Term Memory), ARIMA (Autoregressive Integrated Moving Average), SVM (Support Vector Machine).*

I INTRODUCTION

The stock market refers to the[2] collection of markets and exchanges where regular activities of buying, selling, and issuance of shares of publicly-held companies take place. Such financial events are directed through institutionalized formal exchanges or over the counter marketplaces which operate under a well-defined set of rules. There can be multiple stock trading places in a country or a region which allow trades in stocks and other procedures of securities. So the buying and selling of shares are at high risk associated with it, so various predictions are done to gain more profit and avoid loss. Investors are one type of stakeholders, who may include anyone who has a direct or indirect equity interest in the business or someone with a non-equity interest in a non-profit organization. Shareholders are granted special privileges depending on the class of stock, including the right to vote on matters such as elections to the board of directors based on the stock types, the right to share in distributions of the company's income, the right to purchase new shares issued by the company.

The stock price is an important feature of the finance market. The investors are not sure of shares where buy or sell, they have to predict the prince and take a risk. The prediction of the further or present price is done based on the previous date share price. The price rate of share market and each share are not stable it is continuously changing. The price of shares fluctuates based on many factors like supply and demand, Emergency needs, Economy, politics, etc. The shareholders invest at risk and think twice before it has done. The forecasting accuracy of the share market is directly related to investment and profits.

The stock market is the largest platform for investment. There is mainly two stock exchange in India NSE (NATIONAL STOCK EXCHANGE) Nifty is an index of NSE, BSE (BOMBAY STOCK EXCHANGE) Sensex is an index of BSE. When companies want to list down to share market than the first list to the primary market than in secondary market investors can buy or sell shares of the particular company during the initial public offering. The most important input factors of the company that have an impact on prediction are Price-To-Earnings Ratio (PE), Debt-To-Equity Ratio and Price Flow Ratio (PF)[4]. Getting high profits and without incurring any risk is the main agenda of the investors. Stock prices are highly unstable and very uncertain. In the past, many predictive models have been developed for such purpose and applied to financial market prediction. Commonly they are categorized by

(1) Fundamental analysis
(2) Technical analysis
(3) Time series analysis

Fundamental analysis is[5] a technique of estimating securities by attempting to size the fundamental value of a stock. The study of fundamental analysts the whole thing from the industrial environments and inclusive economy to the financial condition and managing of companies. Earnings, expenditures, assets, and liabilities are all essential features to fundamental analysts. While Technical analysis varies from fundamental analysis in that the stock's amount and volume are the single inputs. The core assumption is that all known fundamentals are factored into the price, thus there is no need to pay close attention to them. Technical analysts do not attempt to measure a security's intrinsic value, but, instead, use stock charts to identify patterns and trends that suggest what a stock will do in the future. Time series analysis involves techniques for analysing time series data in order to extract important information and other features of the data. To predict the future value according to it's previously observed values using the model of time-series Forecasting.

II SURVEY ON EXISTING METHODS

The stocks are considered as fluid and unstable in nature. Still, a vast population invests in stocks day-to-day aiming to return profits from the market. To reduce the false forecasts of the stock market and increase the ability to predict market

movements, and to avoid the risk and the complexity in predicting stock price. Some prediction techniques like LSTM, SVM, ARIMA, RNN (Recurrent Neural Network), CNN (Convolutional Neural Network), and Linear Regression is used to solve this problem.

2.1LSTM (Long Short Term Memory

Long short-term memory (LSTM) is the same as the recurrent neural network (RNN) architecture. It is used in the field of deep learning, unlike standard feed-forward neural networks, LSTM has feedback networks. It can not only procedure single data points (such as images) but similarly entire sequences of data (such as speech or video). For example, LSTM is applicable to tasks such as unsegmented, connected Handwriting Recognition, Music Generation, Language Translation, and Image Captioning[6].
LSTM[8] has a new structure called a memory cell. A memory cell contains four main elements: a forget gate, an input gate, an output gate and a neuron unit which is a central linear unit with a fixed self-recurrent connection. The different gates and their roles are as follows:

- Forget gate layer: it manages how much forget to the previous state to next cell state.
- Input gate: Input gate contains the input.
- Output gate: It contains of the output produced by the LSTM.
- Cell State: Runs through the whole system and has the capacity to add or eliminate info with the help of gates.
- Sigmoid layer makes numbers between zero and one, describing how much of each module should be let through.
- Tanh layer generates a new vector, which is added in the state.

The cell state is modernized based on the productions form the gates. Mathematically we can represent it using the following equations.

- ft= σ (wf .[ht-1 ,xt] + bf) (1)
- it= σ (wi .[ht-1 , xt] + bi) (2)
- ot= σ (wo. [ht-1 ,xt) + bo) (3)
- ct= tanh(wc .[ht-1 , xt) + bc) (4)
- ht= ot * tanh(ct) (5)

Whereft: forget gate vector, xt: input vector, ht: output vector, it: input gate vector, ot: output gate vector, ct: cell state vector, and w, b are the parameter matrix and vector.

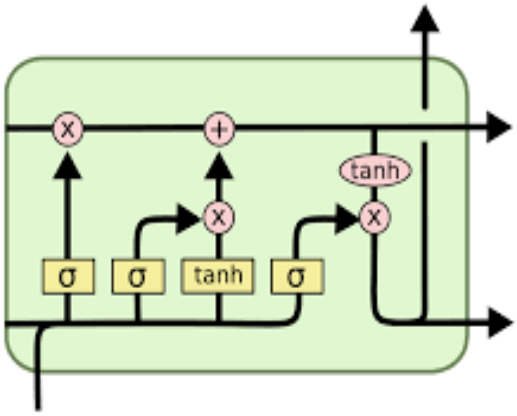

Figure 1: LSTM Architecture[7]

2.2 ARIMA (Autoregressive Integrated Moving Average)

ARIMA stands for an autoregressive integrated moving average. It's a method[9] for predicting time series data for forecasting (i.e., for predicting future points in the series). More precisely, the method involved considering a value Y at time point t and adding/subtracting based on the Y values at previous time points (e.g., t-1, t-2, etc.), and also adding/subtracting error from previous time points.
The formula itself looks like this:

$$Yt = c+\phi 1ydt-1+\phi pydt-p+...+\theta 1et-1+\theta qet-q+et \quad (6)$$

Where "e" is a fault term and "c" is a constant. ARIMA models are expressed like "ARIMA (p,d,q)", with the three terms p, d, and q distinct as follows:

- p means the number of preceding Y values that have to be added/subtracted to Y in the model.
- ddenotes the number of times that the data have to be "differenced" to produce a stationary signal.
- q denotes the number of preceding/lagged values for the error term that are added/subtracted to Y.

2.3 SVM (Support Vector Machine)

The support vector machine (SVM) is a training algorithm for learning classification and regression rules from data, for example[11] the SVM can be used to learn polynomial, radial basis function (RBF) and multi-layer Perception (MLP) classifiers. SVMs were first recommended by Vapnik in the 1960s for organization and have freshly become an area of intense study owing to progress in the methods and theory coupled with extensions to regression and density estimation. SVMs arose from algebraic learning theory; the goal is to resolve only the problematic of interest without solving a more challenging problem as an intermediary step. SVMs are based on the organizational risk minimization principle, closely related to regularization theory. Two key elements in the application of SVM are the methods of mathematical programming and kernel tasks. The factors are found by solving a quadratic programming problem with linear equality and inequality limitations; rather than by solving a nonconvex, unconstrained optimization problem. SVM can be used for mutual classification and regression job. In a classification case, we try to catch an optimal hyperplane that splits two classes. In direction to find an optimal hyperplane, we need to minimalize the norm of the vector w, which describes the splitting hyperplane. This is equal to maximizing the margin among two classes. Mathematically, we will gain a quadratic programming problem where the total number of variables is equal to the total number of observations.

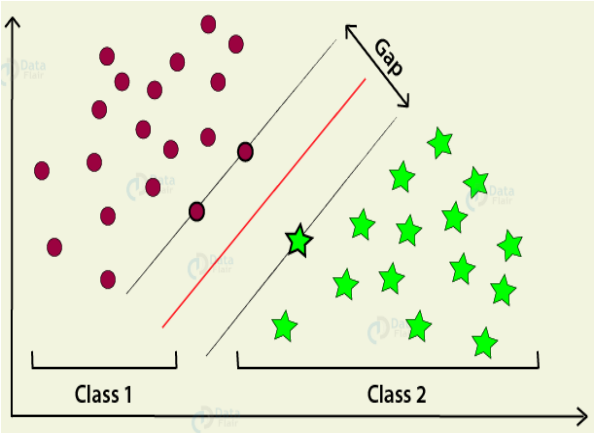

Figure 2: Support Vector Machine[10]

III EXISTING RESEARCH ON STOCK MARKET

Jai Jagwani, et al. (2018) [1], To identify the relation between different existing time series algorithms namely ARIMA and Holt Winter and the stock prices is the main objective of the proposed work, for the investments a good risk-free range of stock prices are analyzed and therefore better accuracy of model can be seen. To find distinguished results for shares in the stock market, the combination of two different time series analysis models is opted by producing a range of prices to the consumer of the stocks. Not complex in nature and estimation of values which are purely based on the past stock prices for non-seasonal or seasonal is the main advantage of these models. In this experiment, some limitations are the work that never takes into consideration and other circumstances like news about any new market strategy or media release relevant to any company which may get affected by the prices of stocks.

R. YaminiNivetha, et al. (2017) [13], A relative study of the three algorithms namely - Multiple Linear Regression (MLR), Support Vector Machine (SVM) and Artificial Neural Network (ANN) is the main aim of this study. To predict the coming day market price, the prediction will be determined by monthly prediction and daily prediction. Sentiment analysis with the best prediction algorithm forecast the stock price. The less-developed algorithm is the Multiple Linear Regression algorithm which calculates the correlation between volume and the stock price. The result of the study shows that deep learning algorithms are more developed than MLR algorithms and SVM algorithm.

Ishita Parmar, et al. (2018) [14], this paper studies the use of Regression and LSTM based Machine learning to forecast stock prices. Factors measured are open, close, low, high and volume. This paper was an attempt to determine the future prices of the stocks of a company with improved accuracy and reliability using machine learning techniques. LSTM algorithm resulted in a positive outcome with more accuracy in predicting stock prices.

Jeevan B, et al. (2018) [15], This paper is mostly based on the approach of predicting the share price using Long Short Term Memory (LSTM) and Recurrent Neural Networks (RNN) to forecast the stock value on NSE data using various factors such as current market price, price-earning ratio, base value and other anonymous events. The efficiency of the model is analysed by comparing the true data and the predicted data using an RNN graph. Machine learning to predict stock price as see the model is able to predict the stock price very close to the actual price where this model captures the detailed feature and uses different strategies to make a prediction. The model train for all the NSE data from the internet and recognize the input and group them and provide input according to the user configuration this RNN based architecture proved very efficient in forecasting the stock price by changing the configuration accordingly which also use backpropagation mechanism while gathering and grouping data to avoid mixing of data.

Youxun Lei, et al. (2018) [16], In this paper, multi-category news events are used as features to develop stock price trend prediction, model. The multi-category events are based on already defined feature word dictionary. And we have

employed both neural networks and SVM models to analyse the relationship between stock price movements and specific multi-category news. Experimental results showed that the predefined multi-category news events are more improved than the baseline bag-of-words feature to predict stock price trend. As compared to long term prediction, short term prediction is better based on this study.

MehakUsmani, et al. (2019) [17], The prominent aim of this study is to forecast the market performance of Karachi Stock Exchange (KSE) on day closing using various machine learning algorithms. The prediction model uses a number of attributes as an input and predicts market as Positive & Negative. The features used in the model include Oil rates, Gold & Silver rates, Interest rate, Foreign Exchange (FEX) rate, NEWS and social media feed. The machine learning algorithms including Single Layer Perceptron (SLP), Multi-Layer Perceptron (MLP), Radial Basis Function (RBF) and Support Vector Machine (SVM) are compared. The algorithm MLP that is multi-layer perceptron performed best as compared to other techniques. The most useful feature in predicting the market was the oil rate attribute. The end results of this research confirm that machine learning techniques have the ability to predict the stock market performance. Multi-Layer Perceptron algorithm of machine learning is a predicted 77% correct market performance.

Tingwei Gao, et al (2017) [18], To assess the scheme that merges RNNs with informative input variables which can give an improved and effective method to forecast the next-day market is the main objective of this paper. The stock prediction model analyses using long-short memory (LSTM) and stock basic trading data. On Standard & Poor's (S&P500) and NASDAQ, the case study relies. The stock closing price is more precisely predicted using their forecasting system for the next day, which outperforms the comparison models. This is the main discovery of the case study. Five various models namely – moving average (MA), exponential moving average (EMA), support vector machine (SVM) and LSTM are tested by them to demonstrate the utility of the system. The closing value of the next day is the predicting target.

Shao-En Gao1, et al (2018) [19], The processed data of past stocks for forecasting the future stock price as well as a deep learning method is applied to find approximate trending value of stock prices which are mentioned in this paper. This paper not only have the data of historical stock with the time scale but also predicts prices of the future stock by a designed neural network, this is due to the fact that the trend of stocks is usually connected to the previous information of stock price. In this paper, the design of the neural network proposed then with the memory performance the convolutional recurrent neural network (CRNN) and for improving the long-term dependency of traditional RNN the Long Short-term memory (LSTM) are the major components. Also to enhance the accuracy as well as stability of prediction of the RNN -LSTM architecture is put. This paper accumulates a total of ten stock historic data to test and accomplish an average error rate of 3.449 RMSE.

Jingyi Du, et al (2019) [20], The LSTM neural network is used to predict Apple stocks by using single feature input variables and multi-feature input variables to verify the prediction effect of the model on stock time series. The testing results show that the model has a high accuracy of 0.033 for the multivariate input and is accurate, which is in line with the actual demand. For univariate feature input, the predicted absolute error is 0.155, which is lesser to the multi-feature variable input.

IV CONCLUSION

In this paper, a review on various stock price prediction techniques has been presented. On the basis of published and existing literature, it can be concluded that the some existing techniques are not appropriate for prediction of stock market trends as well as price of different socks.There exist a gap between technologies and user requirement for an accurate stock prediction system. This paper survey on various machine learning technique such as LSTM (Long Short Term Memory), ARIMA (Autoregressive Integrated Moving Average), SVM (Support Vector Machine) are used to predict target values. If regression and classification learning model is used then SVM technique is best whereas, long term or short term dependencies are available then LSTM technique is best. If time series analysis is used then all three technique is best to predict the time series data. In our case, the target value is the stock price of next day. Among these it can be said that out of this three algorithms LSTM and SVM gives better prediction model.

Method	**Result**		
ARIMA[1]	High Accuracy		
ANN [14]	High Accuracy		
LSTM[15]	Train Score - 0.00106 MSE (0.03 RMSE) Test Score - 0.00875 MSE (0.09 RMSE).		
LSTM[16]	Max epoch 100, Batch Size 64, Learning Rate 0.001		
SVM[17]	Days	BOW-SVM	Event-SVM
	1	57.28%	66.06%
	3	53.10%	62.18%
	5	56.31%	62.09%

	10	57.28%	65.21%	
MLP[18]	77%			
LSTM[19]	Days	100	200	500
	MAE	11.4091	12.2613	14.7709
	RMSE	16.4717	16.7404	20.4668
	MAPE	0.5360	0.5940	0.7240
	AMAPE	0.5311	0.5871	0.7145
LSTM[20]	3.449 RMSE.			
LSTM[21]	Average absolute error – 0.155			

Table1: Result[1][13][14][15][16][17][18][19][20]

REFERENCES

[1] J. Jagwani, M. Gupta, H. Sachdeva, and A. Singhal, "Stock Price Forecasting Using Data from Yahoo Finance and Analysing Seasonal and Nonseasonal Trend," in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2018, pp. 462–467, doi: 10.1109/ICCONS.2018.8663035.

[2] "Stocks: So What Exactly Are Stocks, Anyway? - TheStreet." [Online]. Available: https://www.thestreet.com/topic/1584/stocks.html. [Accessed: 05-Jan-2020].

[3] "How the stock markets work." [Online]. Available: https://cleartax.in/s/stock-market-explained. [Accessed: 05-Jan-2020].

[4] S. E. Gao, B. S. Lin, and C.-M. Wang, "Share Price Trend Prediction Using CRNN with LSTM Structure," in *2018 International Symposium on Computer, Consumer and Control (IS3C)*, 2018, pp. 10–13, doi: 10.1109/IS3C.2018.00012.

[5] "Technical Analysis Definition." [Online]. Available: https://www.investopedia.com/terms/t/technicalanalysis.asp. [Accessed: 05-Jan-2020].

[6] "The magic of LSTM neural networks - DataThings - Medium." [Online]. Available: https://medium.com/datathings/the-magic-of-lstm-neural-networks-6775e8b540cd. [Accessed: 05-Jan-2020].

[7] "Understanding LSTM Networks -- colah's blog." [Online]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/. [Accessed: 27-Dec-2019].

[8] "Stock price prediction using LSTM, RNN and CNN-sliding window model - IEEE Conference Publication." [Online]. Available: https://ieeexplore.ieee.org/document/8126078. [Accessed: 27-Dec-2019].

[9] "Chapter 23: Using ARIMA for Time Series Analysis." [Online]. Available: https://ademos.people.uic.edu/Chapter23.html. [Accessed: 30-Dec-2019].

[10] "Support Vector Machines Tutorial - Learn to implement SVM in Python - DataFlair." [Online]. Available: https://data-flair.training/blogs/svm-support-vector-machine-tutorial/. [Accessed: 07-Jan-2020].

[11] S. S. Patil, K. Patidar, and A. M. Jain, "A Survey on Stock Market Prediction Using SVM," vol. 02, no. 01, p. 7, 2016.

[12] Z. Yeze and W. Yiying, "Stock Price Prediction Based on Information Entropy and Artificial Neural Network," in *2019 5th International Conference on Information Management (ICIM)*, Cambridge, United Kingdom, 2019, pp. 248–251, doi: 10.1109/INFOMAN.2019.8714662.

[13] R. Y. Nivetha and C. Dhaya, "Developing a Prediction Model for Stock Analysis," in *2017 International Conference on Technical Advancements in Computers and Communications (ICTACC)*, Melmaurvathur, India, 2017, pp. 1–3, doi: 10.1109/ICTACC.2017.11.

[14] S. Agrawal, D. Thakkar, D. Soni, K. Bhimani, and C. Patel, "Stock Market Prediction using Machine Learning Techniques," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, pp. 1099–1103, Apr. 2019, doi: 10.32628/CSEIT1952296.

[15] B. Jeevan, E. Naresh, B. P. V. kumar, and P. Kambli, "Share Price Prediction using Machine Learning Technique," in *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*, Bangalore, India, 2018, pp. 1–4, doi: 10.1109/CIMCA.2018.8739647.

[16] Y. Lei, K. Zhou, and Y. Liu, "Multi-Category Events Driven Stock Price Trends Prediction," in *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*, Nanjing, China, 2018, pp. 497–501, doi: 10.1109/CCIS.2018.8691392.

[17] M. Usmani, S. H. Adil, K. Raza, and S. S. A. Ali, "Stock market prediction using machine learning techniques," in *2016 3rd International Conference on computer and Information Sciences (ICCOINS)*, 2016, pp. 322–327.

[19] T. Gao, Y. Chai, and Y. Liu, "Applying long short term momory neural networks for predicting stock closing price," in *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, 2017, pp. 575–578, doi: 10.1109/ICSESS.2017.8342981.

[19] S. E. Gao, B. S. Lin, and C.-M. Wang, "Share Price Trend Prediction Using CRNN with LSTM Structure," in *2018 International Symposium on Computer, Consumer and Control (IS3C)*, Taichung, Taiwan, 2018, pp. 10–13, doi: 10.1109/IS3C.2018.00012.

[20] J. Du, Q. Liu, K. Chen, and J. Wang, "Forecasting stock prices in two ways based on LSTM neural network," in *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, 2019, pp. 1083–1086, doi: 10.1109/ITNEC.2019.8729026.

