ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 11, November 2022

Review Paper on Machine Learning Techniques Used For Prediction of Stock Price

[1] Anany Pandey, [2] Manish Pandey

[1] [2] Computer Science and Engineering, Maulana Azad National Institute of Technology, Bhopal, Madhya Pradesh, India.
Corresponding Author Email: [1] pandeyanany@gmail.com, [2] contactmanishpandey@yahoo.co.in

Abstract*— A marketplace where buyers and sellers meet to trade is called a stock market. A stock market is also known as an equity market, a share bazaar, or a share market. It has been found that many people have become billionaire by trading. Stock price prediction plays a vital role in knowing which stock to hold/sell. Mostly trader have different analysis for predicting stock. Stock market add a lot to GDP of a country. Country like India has lot of potential as it is a developing country. Growth rate of India is quite high as compared to other. Thanks to ML techniques which are helping to get accurate stock prediction. This paper gives insight of the research paper published related to this paper.

Keywords: Stock Market, Machine learning, Artificial intelligence, Artificial Neural networks, technical analysis, Fundamental analysis.

I. INTRODUCTION

The Indian stock market is divided mainly into National Stock Exchange (NSE) and the Bombay Stock Exchange (BSE). Founded in 1875 by a group of traders, is Asia oldest stock exchange is Bombay Stock Exchange (BSE). The first mass participation from public came when Mr. Dhirubai Ambani came up with Reliance IPO in 1980.There are many indices under Bombay Stock Exchange like SENSEX, BSE 100, etc. Established in 1992, the National Stock Exchange has since evolved into an advanced, automated, electronic system that offers trading opportunities to investors across the country. There are many indices under National Stock Exchange like NIFTY, NIFTY 50, etc.

Stock Market gives promising return on investment if analysis of stock is correct. Growing demand of technical analysis has led to advancement of Artificial Intelligence which further has led to advancement of machine learning and deep learning techniques. Even though huge number of investors are present but only fraction of them can make profit. That is why analysis of market is in demand. Now as time is passing by amount of data for processing is also increasing drastically and to do it manually is not a suitable solution.

Further in this paper we will discuss about different types of analysis in-depth. We will also study significant work done on technical analysis.

II. OVERVIEW ABOUT STOCK

A Stock Market is a trading platform where exchange of publicly hold share of company takes place between seller and buyer. SEBI regulates these exchanges of shares and finds out if any illegal trading takes place.

Stocks are broadly classified into 6 types.

a) Large, Well Established, stable stock with excellent reputation is known as Blue Chip stock. Some examples of stocks present in Indian stock market are Reliance, Tata Consumer Service, HDFC Bank, etc. This type of stock is less volatile as compared to other. Here technical analysis plays a vital because the amount of data to be processed are more as they are being trading from long back.

b) Stock that is trading below their below their intrinsic value is known as Undervalued Stock. Some examples these stocks in Indian Stock Market are ITC, ACC, etc. In this to know if the stock is undervalued or not is important.

c) Stocks, with rapidly growing revenue earnings come under Growth stocks. To know when the stock will surge and how much the stock will surge is important. Example of such stock in Indian stock market are Adani, D-MART, Bajaj Finance, etc.

d) Companies that pay regular, High dividend to shareholders is dividend stocks. These stocks are also less volatile as compared to other type of stocks. Mostly companies give dividend on quarterly basis. Some examples of such stocks in Indian stock market are Indian Oil Corporation Limited, Coal India Limited, etc. These stocks are for holding purpose. Analysis of stocks is not that important here.

e) Stock that offers products that people always need and are recession proof is known as defensive stock. Such stocks are most stable stocks. Some of such stock examples are Dabur, Hindustan Uniliver Limited. Analysis is not so important as compared to other type of stocks.

f) Stock which are new into market. Value of such stocks

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 11, November 2022

are very less as compared to other stocks. Example of such is Trident, Solana, etc. As the data available is very less these stocks are very risky but as quoted in stock market "High risk high gain". This are the most promising stocks if analyzed properly. Here analysis is very important as compared to other.

Types Of Analysis Used in Stock Market.

a) Technical analysis is used to predict the gain or loss of prices by studying past market data, mainly price and market cap. Total of 7462 companies are cumulatively present in Bombay Stock Exchange and National Stock Exchange leading to lot of data for processing. Demand for algorithm of technical analysis has rigorously gone up leading to research of more accurate machine learning algorithm. With the right data and correct configuration, an artificial algorithm can capture, and model all input and output relationships. Their accuracy and predictive abilities can be tested both mathematically and empirically by splitting the dataset into a testing and a training dataset in a perfect ratio.

b) Fundamental analysis [19] is used to predict the gain or loss of prices by examining earning, dividend, assets, quality, ratio, new products. Charts are most widely used for fundamental analysis. Fundamental analysis requires human accuracy which is not on level of technical analysis accuracy. Here mostly profits are made by buying wrongly priced stock at low price and wait for market correction to take place. Most famous strategy used by fundamental analyst are to hold and wait.

Participants of Stock Market

a) Stockbroker [12]: Seller and buyers of stock use stockbroker as interface for trade share at a correct market price, leading to creating liquidity in the market. Stockbroker can work as both freelancer and as an employee. Stockbrokers earn by charging some percentage on profit which is made on invested money. Below are the things a person needs to do as a broker.

- Help and advice investors
- Research
- Set Investment Goals
- Take Investment Decision

b) Investment Banker [12]: Person how works as an individual for financial institution that is in the business primarily for raising capital for companies, government, and other entities. Basically a investment banker represents the owner of company.

c) Portfolio Manager [12]: Manges other investors fund by investing them in capital markets. This manger is mostly found in Asset management companies or mutual fund companies or Investment bankers.

Disadvantage of fundamental analysis

a) Lot of data is to be processed and to extract important data and to process becomes difficult.

b) It becomes harder to manage analysis and market at given instant of time.

Stock market indices.

Stock market indices is created using Market-Cap Weightage and Price Weightage. Some of the popular stock market indices are:

a) Benchmark Indices [12]: The collection of stocks is called benchmark indices because they use best practices to regulate the companies they choose.

b) Sectoral Indices [12]: Both BSE and NSE have some good indicators measuring companies that fall into a specific sector. Indices such as S&P BSE Healthcare and NSE Pharma are considered good indicators.

c) Market-Cap Based Indices [12]: Few indices select companies based on their market capitalization. Market value of a listed company can be known through Market Capitalization. Examples of small cap companies are S&P BSE and NSE Small Cap 50.

BASICS OF INDICATORS MOSTLY USED IN STOCK MARKET.

- Open Price: Price at which equity trades when an exchange opens for the day.
- Closing/Raw Price: Price of equity for which it was last transacted before market getting closed.
- Market Capitalization: Referred to as market cap. Market cap = Current Share Price * Total Number of Share Outstanding.
- All Time High: Highest value for which equity was traded for. Mostly All-time high is calculated for 52 weeks.
- All Time Low: Lowest value for which equity was traded for. Mostly All-time low is calculated for 52 weeks.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 11, November 2022

III. VARIOUS MACHINE LEARNING ALGORITHM ANALYSIS FOR STOCK PREDICTION

Table 1: Comparison Of different regression models

Regression Algorithm	Performance Parameter	Supervised learning without time series		Format for supervised learning in time series	
		Value for train-split in percentage.	Value for k-fold cross-validation in percentage.	Value for k-fold cross-validation in percentage.	Value for train-split in percentage.
SVM	Accurate (%)	87.35%	87.22	75.48%	89.33%
	Precised (%)	87.57%	87.69%	84.93%	89.33%
	recall	86.92%	87.09%	70.07%	89.27%
	F-1 value	87.21%	87.19%	70.61%	89.27%
Perceptron neural network	accuracy	75.88%	76.99%	75.48%	76.68%
	precision	83.48%	83.70%	84.93%	84.34%
	recall	74.80%	76.36%	70.07%	74.91%
	F-1 value	73.17%	74.67%	70.61%	73.61%
logistic regression	accuracy	86.98%	87.10%	89.45%	89.93%
	precision	86.72%	87.18%	89.72%	89.98%
	recall	87.20%	87.52%	89.28%	89.81%
	F-1 value	86.93%	87.17%	89.40%	89.87%

Mehar Vijh et. al. [1] (2020) used machine learning algorithm Random Forest (RF) and Artificial Neural Network (ANN)[13] for prediction of stock price for other day. 10-year historic dataset from 4/5/2009 to 4/5/2019 of companies namely, JP Morgan, Nike, Johnson, Goldman Sachs, and Pfizer were taken. For minimizing errors on the predicted price, predicted closing prices were given as input to Mean Bias Error (MBE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) . A comparison clearly demonstrates that ANN outperforms RF in stock price prediction on the basis of MAPE, RMSE and MBE values

Alice Zheng et. al. [2] (2017) in his paper used dataset from NYSE stocks. Randomly 82 stocks were taken. Total of 2,60,000 data for sample was taken. Alpha Vantage API [14] provided him access to time series for intraday, daily weekly, monthly data. Prediction was made for only short term. Daily time series was taken into consideration, which included open price for a day, high price for a day, lowest price for a day, closing price for a day, and volume traded for a given day. Training set and Testing set were partitioned into 9:1 ratio. Classification was done for knowing if price will go up or down. For resolving classification problem, algorithm such as Logistic Regression, Simple Neural Network, Bayesian Network and SVM with RBF kernel in the sklearn library and ran them on the prices of one specific stock named "MSFT" In this paper comparison between Linear Regression and Support Vector Machine was made. Here two metrics: The prediction error rate and mean squared error was used. It was found that SVM gave higher variance for smaller training set as compared to Linear regression. SVM outperformed Linear Regression. Correct rate of 69.5% was achieved and lowest of 68.5% was recorded.

V Kranthi Sai Reddy et. al. [3] (2018) in his paper used Support Vector Machine algorithm for predicting the price of stock. Support Vector Machine (SVM) [17] was considered to be most suitable algorithms for the time series prediction. The algorithm was used for both, regression, and classification. RBF kernel function was used in support vector machine classification. For predicting stock price parameter such as sector momentum, price volatility, sector volatility, price momentum was taken into consideration. Dataset was downloaded from net. log2c and log2g value for minimizing error were used. Numerical result suggested the high efficiency.

Nitin Nandakumar Sakhare et. al. [4] (2019) performed Polynomial regression, Linear regression [21], and Support Vector Regression for predicting price of stock. For dataset S&P 500 was used. Dataset provided updated price of 500 stocks every 60 seconds. 80% of dataset was used for training while 20% was used testing. Average Standard Deviation and Mean Square Error are taken into consideration for measuring prediction accuracy. Out of all prediction technique it was found that Support Vector Regression

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 11, November 2022

training was easy, and irrespective of size of feature known as dimension space it performed quite accurately. Error can also be managed unambiguously.

Irfan Ramzan Parray et. al. [5] (2020) used Perceptron neural network, Support vector machine and logistic regression was used to predict stock price. For dataset 50 stocks from NIFTY 50 was taken. Data was collected from January 1, 2013, to December 31, 2018. Technical indicators like Moving average convergence divergence (MACD), Exponential moving average (EMA), Relative strength index (RSI), Average true rage (ATR), Market Opening Price, today Highest, today lowest, Market closing price, Previous Close price was taken into consideration. 30% was used for testing purpose and 70% of data was used for training. Each algorithm was implemented for supervised learning using non time series format and supervised learning using time series format. Further for both Train–test split value (%) and k-fold cross-validation value (%) was recorded for supervised learning for non-time series and for format of time series. Support Vector Machine reported accuracy of 87.35%, Logistic regression reported accuracy of 75.88% and Perceptron reported accuracy of 76.88%. Further performance detail is given in Table 1.

Akshar Rastogi et. al. [6] (2021) preformed stock market prediction with Lasso regression using Technical Analysis and time lag. Dataset used was of NIFTY 50. Dataset chosen was from 1 January 2009 to December 30, 2020, total of 11 years and 11 months. Then Lasso regression [16] was applied onto it. Time lags used in models are for 1 to 10 days. Newly created dataset was used. Each sample collected contained Opening price of a day, Highest price of a day, Lowest price of a day, Closing price of a day before processing data. Dataset was further divided into 30% testing and 70% training. Accuracy of model was calculated using Mean Absolute Percentage Error and Root Mean Square Error and. To overcome with the issue of Not a Number it was either replaced by mean or dropping its value associated row. It was found that replacing Not a Number value with mean gave better result.

Jibing Gong et. al. [7] (2009) in his predicted stock price using Logistic Regression [20] Model. For dataset Integrated Stock Index data was selected for testing and training samples. Data was taken from 2005 to 2007. Dataset of year 2005 was fully used for training model. Of the 12 months of 2005, only 8 months proved useful. Parameters such as previous closing price for a day, opening price for a day, highest price for a day, lowest price for a day, closing price for a day, composite weight price, daily turnover, amount traded or volume traded for a day, and traded number of a stock were taken into consideration. Only 8 months were taken into because some lack required parameters. Next step was to select optimizing regression coefficient for financial data of year 2006. Average model among all the model was taken into consideration as it was best satisfying. Accuracy of model was found to be 83.3%.

Bui Thanh Khoa et. al. [8] (2022) in his research for predicting prices of stocks using machine learning algorithm made comparison between predictive ability of three forecasting models, including logistic regression, Artificial neural network (ANN) and Support vector machine (SVM). Research was done for short-term investment. For dataset all 30 stocks were chosen from VN30 list. Data collecting period was from July 28,2000 to July30,2021. Date ticker, highest price of a day, opening price closing price of a day of a day, lowest price of a day, and trading volume of a day were indicators taken into consideration from dataset. Because large number of tickers are present in the VN30. So, MWG (Mobile World Investment Joint Stock Company) which is in the list of VN30 was taken as a dataset. For result it was found that logistic regression gave accuracy of 53.84%, Artificial neural network gave accuracy of 60.03% and Support vector machine gave accuracy of 91.52%. It can be seen SVM outperformed logistic regression and Artificial neural network.

Ashish Pathak et. al. [9] (2019) in his research using machine learning and sentiment analysis predicted Indian Stock Market. Approach was to get output accurate recommendation in a quantifiable manner. Historic data was given input to Machine Learning Module and News headline were given to sentiment analysis. Machine learning module gave output as Stock Prediction Value while Sentiment Analysis gave output as Sentiment value which again was given input to Fuggy Logic Module. Fuzzy Logic Module gave us the required stock Faith. For machine learning module indicators such as opening price of the day, day lowest and day highest were taken, Exponential moving average, simple moving average of the day's stock first time trading and stock last time trade prices, exponential moving average of the day's lowest and highest prices, Bollinger Bands of the day's opening and closing prices, and Bollinger Bands of the day's lowest and highest prices. For machine learning module ridge regressor [15] was used. It was selected on experimental basis. For sentiment analysis data was collected from www.moneycontrol.com. Then that news went through process of Tokenizing, Lemmatizing, and finding most informative features. Given feature were then classified into positive or negative. Fuzzy computed input from machine learning module and sentiment analyzer module and accurate output.

Mruga Gurjar et. al. [10] (2018) in his paper of stock prediction using Artificial Neural Network. For predicting price of stock indicators such as Stochastic Oscillator, Moving Averages, Standard Deviation, On-Balance Volume was taken into consideration. For dataset Nifty50 stocks were considered and stock was predicted for nest day, next 3 day, next 5 day. Dataset was divided into training and testing

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 11, November 2022

dataset. ANN with backpropagation was used to predict price stock. Forward propagation in neural network was done during training phase to get accurate result. While feeding input to ANN forward, inputs are feeded to neurons. Formulae for TI (Total input of the neuron) was calculated. Optimization method gradient descent was used for training dataset.

$TI= N_1*W_1 + N_2*W_2 + N_3*W_3\ldots\ldots N_x*W_y + 1*W_b$

$N_1, N_2\ldots\ldots N_n$: Represents input neurons.

$W_1, W_2\ldots\ldots W_y$: Represents input neurons weight.

W_b : Weight associated with bias.

Output Of Neuron: $(1+e^{-TI})$.

Gourav Bathla et. al. [11] (2020) in his paper compared Long Short Term Memory and Support vector regression over dataset of S&P 500, NSE (National Stock Exchange), NYSE (New York Stock Exchange), BSE (Bombay Stock Exchange), NASDAQ and Dow Jones industrial Average for experiment analysis. RNN's (Recurrent Neural Network) modified version is Long Short Term Memory. Regression relies heavily on the choice of kernel function, which can be either linear, RBF, sigmoid, or polynomial. To increase the prediction accuracy, the Minimum Loss function is used. The dataset was ingested from January 2015 to January 2020, using data from January 2015 to June 2019 for training and data from July/2019 to January/2020 for testing model. It was found that LSTM outperformed SVR.

IV. CONCLUSIONS

Stock Market is seen to be contributing a lot to countries GDP. Stock Market looked up on as an investment platform. People have opportunity to invest in any favorable stock they want to. Stocks are divided into sector to make It easy to invest.

In this we reviewed many available machines learning for stock prediction paper. Some papers were related to comparison between various machine learning algorithm, somewhere about gaining more accuracy in predicting stock price. Dataset also played a vital role in getting the accuracy. Some papers we reviewed were for short term while some were for long term investment.

Though there has been quite accurate algorithm being found. But none of them was found to be fully accurate. With fundamentally analysis and technical analysis together it can be found that greater accuracy can be obtained.

There is lot of future scope in this domain. More research in this work will result into more efficient and accurate algorithm.

REFERENCES

[1] Vijh, M., Chandola, D., Tikkiwal, V.A. and Kumar, A., 2020. Stock closing price prediction using machine learning techniques. Procedia computer science, 167, pp.599-606.

[2] Zheng, A. and Jin, J., 2017. Using ai to make predictions on stock market. cs229. stanford. edu.

[3] Reddy, V.K.S., 2018. Stock market prediction using machine learning. International Research Journal of Engineering and Technology (IRJET), 5(10), pp.1033-1035.

[4] Sakhare, N.N. and Imambi, S.S., 2019. Performance analysis of regression based machine learning techniques for prediction of stock market movement. International Journal of Recent Technology and Engineering, 7(6), pp.655-662.

[5] Parray, I.R., Khurana, S.S., Kumar, M. and Altalbe, A.A., 2020. Time series data analysis of stock price movement using machine learning techniques. Soft Computing, 24(21), pp.16509-16517.

[6] A. Rastogi, A. Qais, A. Saxena and D. Sinha, "Stock Market Prediction with Lasso Regression using Technical Analysis and Time Lag," 2021 6th International Conference for Convergence in Technology (I2CT), 2021, pp. 1-5, doi: 10.1109/I2CT51068.2021.9417935.

[7] Gong, J. and Sun, S., 2009, June. A new approach of stock price prediction based on logistic regression model. In 2009 International Conference on New Trends in Information and Service Science (pp. 1366-1371). IEEE.

[8] Khoa, B.T. and Huynh, T.T., 2022. Forecasting stock price movement direction by machine learning algorithm. International Journal of Electrical and Computer Engineering (IJECE), 12(6), pp.6625-6634.

[9] Pathak, A. and Shetty, N.P., 2019. Indian stock market prediction using machine learning and sentiment analysis. In Computational Intelligence in Data Mining (pp. 595-603). Springer, Singapore.

[10] Gurjar, M., Naik, P., Mujumdar, G. and Vaidya, T., 2018. Stock market prediction using ANN. International Research Journal of Engineering and Technology, 5(3), pp.2758-2761.

[11] Bathla, G., 2020, November. Stock Price prediction using LSTM and SVR. In 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC) (pp. 211-214). IEEE.

[12] Patel, R., Choudhary, V., Saxena, D. and Singh, A.K., 2021, June. REVIEW OF STOCK PREDICTION USING MACHINE LEARNING TECHNIQUES. In 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 840-846). IEEE.

[13] Breiman, L., 2001. Random forests. Machine learning, 45(1), pp.5-32.

[14] Vantage, Alpha., Alpha Vantage - Free Apis for Realtime and Historical Financial Data, Technical Analysis, Charting, And More! Alphavantage.co. N.p., 2017. Web. 19 Nov. 2017.

INTERNET RESOURCES

[15] Regularization Part 1: Ridge (L2) Regression - YouTube

[16] Regularization Part 2: Lasso (L1) Regression - YouTube

[17] Support Vector Machines Part 1 (of 3): Main Ideas!!! - YouTube.

[18] Technical analysis - Wikipedia.

[19] Fundamental analysis - Wikipedia.

[20] StatQuest: Logistic Regression - YouTube.

[21] Linear Regression, Clearly Explained!!! - YouTube