

An Analysis of Prediction Techniques for Stock Market

Asutosh Dube¹, Srivaramangai R²

Department of Information Technology, University of Mumbai, Mumbai, India

Abstract: *Stock Market is highly volatile market as there are buyers and sellers that try to determine the true price of the stock of a given company by buying and selling the stocks. This is done using Technical Analysis and Fundamental Analysis of the company. In today's era of digitization, we are not only using technology to view the stock price but also to determine the future move of that stock. In this paper we present a literature survey of the different techniques/models we found and compared, that determine the future price of the stock. Artificial Neural Network, Deep Learning, Machine Learning are some of the Techniques and models like ARIMA, LSTM, Linear Regression to name a few. We did a comparative study of the achievements and challenges of each model and towards the end we conclude with a model with highest Accuracy.*

Keywords: Prediction Tools, Fundamental – Technical Analysis, Machine Learning, Deep Learning, Neural Network

1. Introduction

Stock Market is considered as highly volatile due to its nature of price movement, the stock price of a public listed company in the stock exchange changes due to the Buyers and Seller of that stock. The process of determining the true value of the stock is done using Analysis of the companies working and periodic reports published by them. Broadly two types of analysis are used to determine the true value of a stock:

- Fundamental Analysis
- Technical Analysis

A. Fundamental Analysis

To determine accurate value of the stock, the investors need a reliable and accurate information on the financial report of the company. This helps them to take their investing decision. "If the calculated value is higher than the market value of the stock, we can invest else it is a bad investment". We use parameters such as book value, earnings, p/e ratio, ROI etc, to calculate the true value. Fundamental analysis is useful in long-term prediction due to their systematic approach and their ability to predict changes.

B. Technical Analysis

Contrary to Fundamental Analysis, Technical Analysis is based on the existing data of the stock's price movement. Investors analyse the constantly changing attributes in response to different factors make stock prices to rise or fall". Various technical factors of quantitative parameters can be used to analyse the stock price, trend indicators, daily highest price and lowest price, indices, volume to name a few. Following the price movements from the past and deriving insights from the data helps Investors to take their investing decisions. Technical Analysis helps channelize the Trend. Technical analysis is useful for long term as well as short term prediction. Technical analysis data is preferable over fundamental analysis data as input to system.

The above-mentioned techniques are traditionally used for Stock price Prediction for ages and they have proved to be useful to many investors. But the disadvantage with these techniques is the "Time" it takes to calculate the price and

the "Knowledge" possessed by the individual about the same. It becomes difficult for a new investor to channelize their investments if they fail to have the knowledge of these techniques.

Ever since the new advancements in the Digital Field, there have been efforts to reduce the manual task involved in predicting the stock price. Artificial Intelligence systems with Machine Learning and Deep Learning models have proven results in the same. Learning algorithms like (i) Logistic Regression, (ii) K-Nearest Neighbor (iii) Decision Tree, (iv) Bagging, (v) Boosting, (vi) Random Forest, (vii) Artificial Neural Network, (viii) Support Vector Machines (ix) Multivariate Regression, (x) Long-and Short-Term Memory Network (xi) Recurrent Neural Network were used for classification and regression purpose. [1] [6].

While referring these techniques it was found that every model gives different result based on the type of input provided. Some models very useful for long term prediction while the others gave good result for short term prediction

2. Overview of Prediction Algorithms

The quest for the most accurate model has always intrigued researchers. This has led to the exploitation of different method that have the ability to predict. In this section we will discuss about the various techniques used for predicting the Future movement/price of a stock. We need to keep in mind that Stock market data are non-linear, non-stationary, time series data. The basic idea is to train the model with historic data and to predict the future by learning for the past.

A. Machine Learning Techniques

Machine Learning is a part of Artificial Intelligence where the model is fed with past data and the model tries to extract information and patterns. Following are some of the models that have been used for Stock Price Prediction.

• Linear Regression

Linear Regression is a supervised predictive model. It is used to predict the price of the stock based on certain

Attributes [15]. In linear regression we define two variables 'x' and 'y' where y is a dependent variable of x. The model thus generates the result in the form of a regression line. The idea behind Linear Regression model is that there exists a linear relation between two variables.

The advantage with this model is that it has a high accuracy when subjected to RMSE Error Reduction Technique. But the disadvantage is that the real-life problems are not always linear in nature. Thus, as the data gets complex the model fails to fit all the possible events leading to problems like Overfitting and Outlier detection.

- **ARIMA (Auto-Regressive Integrated Moving Average)**

ARIMA is a statistical Machine Learning model used for financial Time Series Forecasting for a short-term period of time [7]. The idea of ARIMA is that the future value is the sum of past values and errors in a linear combination.

ARIMA model has the advantage of providing a directional move whether positive or negative and it can find produce results with data as small as 50-100 observations, the model can be customized by changing the input parameters.

The disadvantage with this model is that it can only give the short-term Prediction.

The model tends to be unstable, both with respect to changes in observations and changes in model specification. Additionally, the model fails to produce long-term prediction.

B. Deep Learning Model

Deep Learning is a sub-field of Machine Learning that is inspired by the functioning of the human brain. As the brain consist of Neurons, Axions and Dendrites that allow helps the brain signalling and proper functioning of the body. Similarly, the Deep Learning Model consists of Neurons arranged in a layered format that generate the output. Following are some of the examples of Deep Learning Models.

- **ANN (Artificial Neural Network)**

Artificial Neural Network or ANN is an intelligent model used proactively for data mining and hidden pattern recognition [4]. It is a set of input and output node with a connected weight. It consists of an input layer where the different parameters are feed to the model as weights, the data is then carried to the hidden layer where the activation function is applied. Finally, the output is received at the output layer which is a single node.

ANN has the advantage of working easily with unstructured dataset. It can be used in parallel data processing. ANN has the ability to model complex non-linear relationship. It can store information obtained for the data. The biggest disadvantage of ANN is the problem of Overfitting. Also, high computational processing and time is required.

- **MLP (Multi-Layered Perceptron)**

Multi-Layered Perceptron also known as Feed Forward Network model is a unidirectional Neural Network

architecture that consists of an input layer, hidden layer and an output layer. The concept of MLP is that the input at any layer is the weighted sum of the output from the previous layer and the Activation function. [13]

The result from any node is never feed back to that same node. MLP is a very simple yet effective model. The conversion of this input into the output is performed by a continuous and differentiable activation function. The learning algorithm, generally a kind of gradient descent algorithm, adjusts the weights of the neurons necessary to reduce the error.

MLP comes with advantage that it can be applied to complex non-linear problems and large amount of data. But the disadvantage is that it requires proper feature selection and quality of training. Also, the extent of independent variable affected by the dependent variable is unknown.

- **LSTM (Long-Short Term Memory)**

LSTM is a special type of Recurrent Neural Network with cells and gates to perform classification and prediction on a dataset [3]. It works best for a time series type of data.

The LSTM model consists of memory cells where each cell has three gates namely the input gate, the forget gate and the output gate. The input of any cell is the output from the previous cell and the cell state. LSTM has the capacity to hold the necessary information for a longer duration and discards the unwanted data using the forget gate.

The Advantage of this model is that it has does not require parameter tuning due to its wide range of inputs. The error propagation is minimum. Also solves the problem of Vanishing Gradient using the memory cells.

3. Literature Survey

Hiransha Ma, Gopalakrishnan E. A [2] conducted a comparative study on ARIMA which a linear model with various deep learning models like MLP, CNN, RNN and LSTM. The comparison was done to predict the stock over a duration of 400+ days. From the result obtained, the models were capable of identifying the patterns the stock markets. ARIMA model being a univariate in nature was not capable of identifying underlying dynamics within various time series. Thus, the results confirm that Deep Learning models are outperforming ARIMA model. Also, CNN has performed better than other networks due to the capability of capturing the abrupt changes in the system with a fixed window is used for predicting the next instant. Mehar Vijh, DeekshaChandola, Vinay Anand Tikkiwal [4] in their research compared Artificial Neural Network and Random Forest Method to predict the future value of the stock. They used the following variables for predicting the closing price.

Stock High minus Low price (H-L)
 Stock Close minus Open price (O-C)
 7 days moving average
 14 days moving average
 21 days moving average
 7 days standard deviation

For the models to minimize the margin of errors they with subjected to Error Reduction methods namely Root Mean Squared Error (RMSE), Mean Bias Error (MBE) and Mean Absolute Percentage Error (MAPE). The results from these methods gave a big win to ANN with values 0.42, 0.013 and 0.77 for RMSE, MBE and MAPE respectively. This clearly indicates that ANN provides better prediction over Random Forest. R. Sathishkumar, R. Girivarman, S. Parameswaran, V. Sriram [5] in their research used the stock market data collected from the 'nseindia' website. Sentiment analysis data is collected through web scrapping using feed parse to get the rss (Really Simple Syndication). The final result would be a combination of LSTM result and Sentimental analysis value. The system used for stock price prediction combining the long short-term memory for technical analysis and sentimental analysis for fundamental analysis successfully gave good accuracy. Sentimental analysis was done based on the keyword narrowed down. This system would be useful for the user who is unaware of stock market. People with minimum or optimum trading experience can also use this system for stock price prediction in future. Jaydip Sen [6] & Mahla Nikou, Gholamreza Mansourfar, Jamshid Bagherzadeh in their research used Machine Learning and Deep Learning Techniques for predicting the closing price of the stock. The paper broadly covers two types of methods i. e., Classification and Regression methods for particularly 2 stocks. Classification techniques used were: (i) Logistic Regression, (ii) K-Nearest Neighbor (iii) Decision Tree, (iv) Bagging, (v) Boosting, (vi) Random Forest, (vii) Artificial Neural Network, and (viii) Support Vector Machines. Regression methods that were built are: (i) Multivariate Regression, (ii) Decision Tree, (iii) Bagging, (iv) Boosting, (v) Random Forest, (vi) Artificial Neural Network, (vii) Support Vector Machine, and (viii) Long-and Short-Term Memory Network. The Result shows LSTM performs better in majority of the cases [6] [11]. Madhusudan Reddy*, Arun Gade, Sreekarreddy, P. Prabhu [8] in their research used ANN and Linear Regression model individually. In the research Exchange data were collected by the Standard & Poor's 500 (S & P 500). The Dataset used has a sample size of 5 years. In each sample, the starting price, the highest price, the lowest price, the closing price, the total number of shares traded was the parameters used. The proposed process consists of different steps like data collection, data pre-processing, data classification and model evaluation. As we know that, Linear Regression is a machine learning algorithm based on supervised learning and ANNs are nonlinear statistical models which display a complex relationship between the inputs and outputs to discover a new pattern. It was observed that both the individual models were able to produce near actual prediction having Standard Deviation with a range of -1 to +1.

Ajinkya Rajkar, Aayush Kumaria, Aniket Raut, Nilima Kulkarni [9] proposed a RNN model used to forecast stock market prices in their paper. The model was a web application developed using Django and React. The results from their research sighted that the News regarding a particular Sector/Stock would impact the price directly on the consecutive day in the market. If the news was positive, the stock would give positive movement and vice versa. [9]

Xiao Ding, Yue Zhang, Ting Liu, Junwen Duan in their research [10] proposed a system that could predict the price based on the events that occurred in the news from various sources using a deep learning method. The model would first get the news that is extracted in form of text and then train the model using neural network. Next, they used a deep CNN to verify the influence of the short-term as well as long term impact of the news/events on the price movement in the market. The result concluded that event driven features have an edge over the traditional parameters in price prediction. Deep learning is useful for news/event based stock price movement prediction. By proposing a neural network to learn event embeddings, and using a deep convolutional neural network to model the combined influence of long-term events and short-term events on stock price movement. Experimental results showed that event embeddings-based document representations are better than discrete events-based methods, and deep convolutional neural network can capture longer-term influence of news event than standard feed forward neural network. Ehsan Hoseinzade, Saman Haratizadeh [12] proposed a variation to the traditional CNN model known as 2D CNN and 3D CNN-based framework, that can be applied on a collection of data from a variety of sources, including different markets, in order to extract features for predicting the future of those markets. The input to the 3D-CNNpred is a matrix of 60 by 5 with depth of 82. The first convolutional layer uses eight filters to perform 1x1 convolutional operation, after which there is one convolutional layer with eight 3x5 filters followed by a 2x1 max pooling 420 layer. Then, another convolutional layer utilizes eight 3x1 filters, again followed by a 2x1 max-pooling layer generates the final 104 features. In the end, a fully connected layer converts 104 neurons to 1 neuron to produces the final output. When the model was compared to other baseline CNN models the 2D and 3D-CNN models performed better due to their rich set of variables and feature extracting layers. The model gave 3% – 11% better result compared to the baseline model. The model performed the ANN model as ANN produced generalized results with shallow feature extraction. Kaustubh Khare, Omkar Darekar, Prafull Gupta, Dr. V. Z. Attar [13] presented a research comparing the Feed Forward Neural Networks and Recurrent Neural Networks. Their experiment essentially focuses on the prediction of these short-term prices exploiting the power of technical analysis. MLP has outperformed LSTM model, in predicting short term stock prices. Also due to the selected parameters Trend_type, Oscillator_type, Momentum_type the MLP gave a granular result which was close to actual prices.

Hadi Rezaei, Hamidreza Faaljou, Gholamreza Mansourfar [14] have come up with a hybrid model for better stock price prediction. Recent developments in deep learning and methods such as LSTM and CNN models, significant improvements have been obtained in the analysis of this type of data. Further, Empirical Mode Decomposition (EMD) and Complete Ensemble Empirical Mode Decomposition (CEEMD) algorithms decomposing time series to different frequency spectra are among the methods that could be effective in analyzing financial time series. They proposed a hybrid model i. e., CEEMD-CNN-LSTM and EMD-CNN-LSTM. The individual stand-alone CNN and LSTM models performed poorly as compared to the hybrid models.

Additionally, CEEMD-CNN-LSTM had relative superiority over the EMD-CNN-LSTM algorithm and provided more accurate results. Maheswari P. and Jaya A [15] compared a machine learning algorithm – Linear regression with a deep learning algorithm – LSTM on a dataset of 5 sectors of stock market from the period of 28th April 2020 to 30th July 2020.

- Pharmaceutical sector: Cipla and Torrent Pharma;
- Banking sector: SBI, ICICI;
- Automobile Sector: Maruti Suzuki Ltd and Mahindra & Mahindra;
- FMCG sector: Hindustan Unilever, ITC and
- Power sector: Power Grid Ltd, Adani Power Ltd.

The results are based on the Root Mean Square Error value which clearly showed that LR model has outperformed considering the RSME factor. Also, this could be due to dataset of the pandemic situation that the entire market crashed in a way leading to a down trend. Qihang Ma [16] in his research compared the three models, ARIMA, ANN and LSTM for stock price prediction. The prediction of the ANN model tends to predict the numerical value of the stock price. Also, the obtained result is processed by using Z-score. The ARIMA model's prediction is directional, which comes from its model assumptions are linear. He Concluded that ANN model is better than that of the ARIMA model, and the performance of the LSTM model.

4. Conclusion

After the thorough review of literature, it is observed that LSTM model works the best for Sequential Type of Data. It has the capacity to extract patterns from historic data and predict the next move accurately as compared to other models. LSTM also solve the major problem of Vanishing Gradient, using the forget gate. Also, LSTM holds the ability to retain the information for longer period of time. It is also observed that Hybrid LSTM models yield in higher accuracy as compared to standalone LSTM model.

References

- [1] PolamuriSubba Rao, K. Srinivas & A. Krishna Mohan A Survey on Stock Market Prediction Using Machine Learning Techniques ICDSMLA Year 2019 Edition 1 DOI: 10.1007/978-981-15-1420-3_101 page 923-931
- [2] Hiransha Ma, Gopalakrishnan E. Ab, Vijay Krishna Menonab, Soman K. P NSE Stock Market Prediction Using Deep-Learning Models International Conference on Computational Intelligence and Data Science (ICCIDS) Year 2018 DOI: 10.1016/j.procs.2018.05.05 page 1351–1362
- [3] Jingyi Shen and M. OmairShafq Short-term stock market price trend prediction using a comprehensive deep learning system Journal Big Data Year 2020 DOI: 10.1186/s40537-020-00333-6 page 7-66
- [4] Mehar Vijha, DeekshaChandola, Vinay Anand Tikkiwal, Arun Kumar Stock Closing Price Prediction using Machine Learning Techniques International Conference on Computational Intelligence and Data Science (ICCIDS) Year 2019 DOI: 10.1016/j.procs.2020.03.326 page 599-606
- [5] R. Sathishkumar, R. Girivarman, S. Parameswaran, V. Sriram Stock Price Prediction Using Deep Learning and Sentimental Analysis Journal of Emerging Technologies and Innovative Research (JETIR) May 2020, Volume 7 Issue 5 ISSN-2349-5162 page 346-354
- [6] Jaydip Sen Stock Price Prediction Using Machine Learning and Deep Learning Frameworks Institute of Electrical and Electronics Engineers (IEEE) page 1-9
- [7] Ayodele A. Adebisi., Aderemi O. Adewumi, Charles K. Ayo Stock Price Prediction Using the ARIMA Model International Conference on Computer Modelling and Simulation Year 2014 DOI 10.1109/UKSim.2014.67 page 105-111
- [8] Madhusudan Reddy*, Arun Gade, Sreekarreddy, P. Prabhu Stock Market Price Forecasting by Using Deep Learning International Journal of Engineering &Technology Year 2018 Edition 7 page 627-631
- [9] Ajinkya Rajkar], Aayush Kumaria, Aniket Raut, Nilima Kulkarni Stock Market Price Prediction and Analysis International Journal of Engineering Research & Technology (IJERT) Volume 10 Issue 06 June-2021 ISSN: 2278-0181 page 115-119
- [10] Xiao Ding, Yue Zhang, Ting Liu, Junwen Duan Deep Learning for Event-Driven Stock Prediction International Joint Conference on Artificial Intelligence (IJCAI) Year 2015 page 2327-2333
- [11] Mahla Nikou, GholamrezaMansourfar, Jamshid Bagherzadeh Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms Intell Sys Acc Fin Mgmt.2019 DOI: 10.1002/isaf.1459 page 1-11
- [12] Ehsan Hoseinzade, Saman Haratizadeh CNN-based stock market prediction using a diverse set of variables Expert Systems with Applications Year 2019 Volume 129 DOI: 10.1016/j.eswa.2019.03.029 page 273-285
- [13] Kaustubh Khare, Omkar Darekar, Prafull Gupta, Dr. V. Z. Attar Short Term Stock Price Prediction Using Deep Learning IEEE Year 2017 DOI: 978-1-5090-3704-9/17 page 482-486
- [14] Rezaei, H., Faaljou, H., Mansourfar, G. Stock Price Prediction using Deep Learning and Frequency Decomposition Expert Systems with Applications Year 2020 DOI: 10.1016/j.eswa.2020.114332 page1-26
- [15] Maheswari P. and Jaya A Prediction of the Stock Market Using Machine Learning–Based Data Analytics Machine Learning Approach for Cloud Data Analytics in IoT Year 2021 page 347-374
- [16] Qihang Ma Comparison of ARIMA, ANN and LSTM for Stock Price Prediction ISEESE Year 2020 DOI: 10.1051/e3sconf/202021801026 page 1-5
- [17] Manish Agrawal, Asif Ullah Khan, Piyush Kumar Shukla Stock Indices Price Prediction Based on Technical Indicators using Deep Learning Model IJRTE, Volume-8 Issue-2 Year 2019, DOI: 10.35940/ijrteB3048.078219, page 2297 – 2305
- [18] Manish Agrawal1, Piyush Kumar Shukla2, Rajit Nair3, Anand Nayyar4, 5, and Mehedi Masud6Stock Prediction Based on Technical Indicators Using Deep Learning Model CMC, 2022, vol.70, no.1, DOI: 10.32604/cmc.2022.014637 page 287 – 304

- [19] Chenjie Sang, Massimo Di Pierro, Improving trading technical analysis with TensorFlow Long ShortTerm Memory (LSTM) Neural Network, The Journal of Finance and Data Science 5 (2019) 1e11, doi. org/10.1016/j. jfds.2018.10.003 page 1-11
- [20] Deniz Can Yıldırım¹, Ismail Hakkı Toroslul^{1*} and Ugo Fiore², forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators, Yıldırım et al. FinancInnov (2021), doi. org/10.1186/s40854-020-00220-2
- [21] Prof. Pritam Ahire, ²Hanikumar Lad, ³Smit Parekh, ⁴Saurabh Kabrawala, Lstm Based Stock Price Prediction, 2021 IJCRT | Volume 9, Issue 2 February 2021 | ISSN: 2320-288 page 5118-5122