International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Survey on Stock Market Prediction using Machine Learning Approaches

***Shraddha Jain*[1]*,Megha Gupta*[2]**

M.Tech Scholar (CSE)[1]*, Department of Computer Science and Engineering, PCST, Indore (India)*
Assistant Professor (CSE)[2] *Department of Computer Science and Engineering, PCST, Indore (India)*

A B S T R A C T

Stock Market prediction is a category of time series prediction which extremely challenging due to the dependence of stock prices on several financial, socio-economic and political parameters etc. Moreover, small inaccuracies in stock market price predictions may result in huge losses to firms which use stock market price prediction results for financial analysis and investments. Off late, soft computing techniques are being used widely for stock market prediction due to relatively higher accuracy compared to conventional statistical techniques. The proposed work investigates the necessity of soft computing approaches for stock market prediction, the various soft computing approaches relevant to the area, previous work in the area with their salient features and the major challenges or the problem domain pertaining to the area. The main focus is on neural networks and fuzzy logic for developing a predictive model. Stock market is highly volatile and is unarguably very difficult to predict accurately based on certain parameters. This paper presents a comprehensive survey on the various techniques explored for high accuracy prediction emphasizing on the salient features.

Keywords: Stock Market Prediction, Artificial Intelligence, Artificial Neural Networks, Fuzzy Logic, Accuracy

1. INDRODUCTION

Stock Markets have long remained one of the one avenues on the forefront which is crucial for the operations of most of the top most companies. Several decisions pertaining to investments, shares etc. depend on the behaviour of the stocks of a company. The stock price values are often leveraged by financial and investment firms for gaining profits and investing. However, the volatile nature of the stock markets make it a risky proposition. Therefore, estimating the future trends in stock prices is somewhat mandatory for firms analysing stock prices and aiming to gain leverage. This calls for stock market forecasting or stock market prediction. Stock market prediction is basically a time series prediction problem. Mathematically:

$$\boldsymbol{P} = \boldsymbol{f}(\boldsymbol{t}, \boldsymbol{v}) \quad \textbf{(1)}$$

Here,

P represents stock price f represents a function of t is the time variable

v are other influencing global variables

The dependence of stock process over time makes it somewhat predictable under similar other conditions of global influencing variables. However, even the slightest of changes can derail the prediction completely.

* *Corresponding author.*
E-mail address: jainshraddha1204@gmail.com

2. NEED FOR ARTIFICIAL INTELLIGENCE BASED TECHNIQUES FOR STOCK MARKET PREDICTION

Primarily, artificial intelligence (AI) based techniques are used where the data to be analyzed is extremely large and complex to be analyzed by conventional computational or statistical techniques. There are various AI based approaches used for time series prediction or fitting applications out of which neural networks and fuzzy logic have gained substantial prominence. With the advent of deep learning, the computational capability of algorithms have also risen allowing us to find trends in highly non-linear and uncorrelated data.

The following section briefly explains the fundamentals of neural networks and fuzzy systems and their application to stock market prediction.

Artificial Neural Networks (ANN)

Artificial Neural Networks try to copy or emulate the thinking process of the human brain to predict data. The fundamental properties are:

1) Parallel data processing capability
2) Learning and Adapting capability
3) Self-Organization

The mathematical model of ANN is shown below:

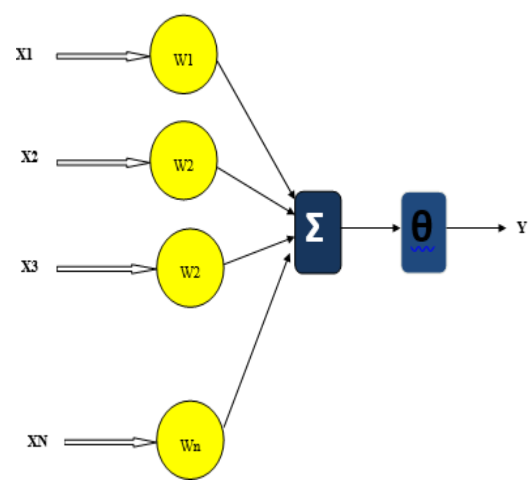

Fig.1 Mathematical Model of Neural Network

The output of the neural network is given by:

$$\sum_{i=1}^{n} XiWi + \Theta \quad (2)$$

here

Xi represents the signals arriving through various paths,

Wi represents the weight corresponding to the various paths and Θ is the bias.

The essence of neural networks lies in the fact that neural networks can find a relation among variables which may seem highly uncorrelated. Moreover, as the data keeps changing, the neural network structure keeps adapting in terms of the weights so as to optimize the model and reduce the errors in the output. Fundamentally, the learning is sub categorized as under:

1) Computational Intelligence
2) Artificial Intelligence
3) Machine Learning
4) Deep Learning

A set theoretic relationship among the above can be seen to be depicted by figure 2

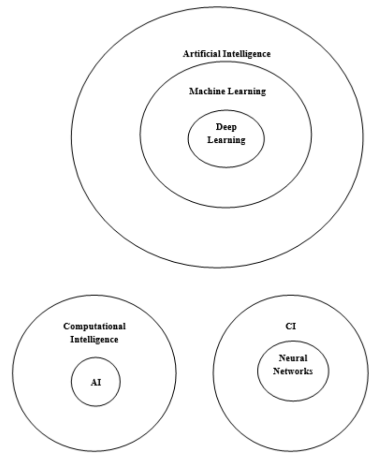

Fig.2 Relationship between machine learning paradigms

Moreover, neural network architectures are also categorized as:

Feedforward Networks: Feed forward networks consist of only the feed forward path for data to travel from input layer towards output layer

Recurrent Networks: Recurrent networks have at least one closed data path loop.

Back Propagation: Back Propagation feeds back the error at the output as an input

The diagrammatic representations are given below:

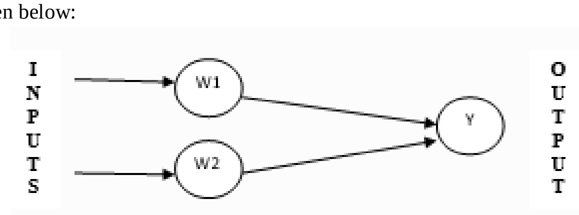

Fig.3 Single Layer Feed Forward Network

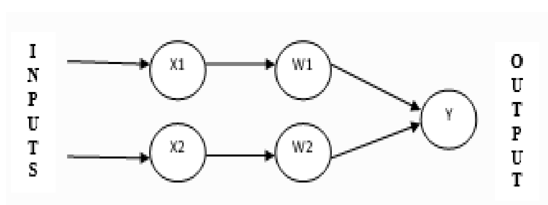

Fig.4 Multi-Layer Feed Forward Network

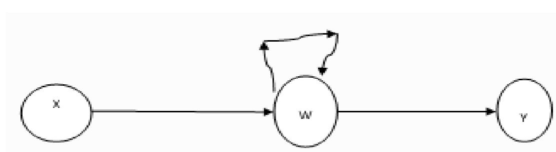

Fig.5 Recurrent Network

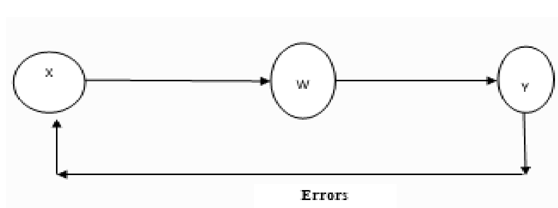

Fig.6 Network with back propagation

Out of the empirical neural network architectures, back propagation has gained significance due to its accuracy in prediction of time series applications. However the performance for different training algorithms vary considerably in terms of errors and number of iterations.

Fuzzy Logic

Another tool that proves to be effective in several prediction problems is fuzzy logic. It is often termed as expert view systems. It is useful for systems where there is no clear boundary among multiple variable groups. The relationship among the inputs and outputs are often expressed as membership functions expressed as:

A membership function for a fuzzy set A on the universe of discourse (Input) X is defined as:

$$\mu A : X \rightarrow [0,1] \quad (3)$$

Here,

each element of X is mapped to a value between 0 and 1. It quantifies the degree of membership of the element in X to the fuzzy set A.

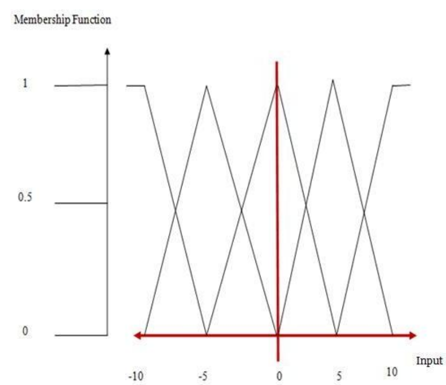

Fig.7 Graphical Representation of Membership Functions

Here,

x axis represents the universe of discourse (Input).

y axis represents the degrees of membership in the [0, 1] interval.

The final category is neuro fuzzy expert systems which governs the defining range of the membership functions.

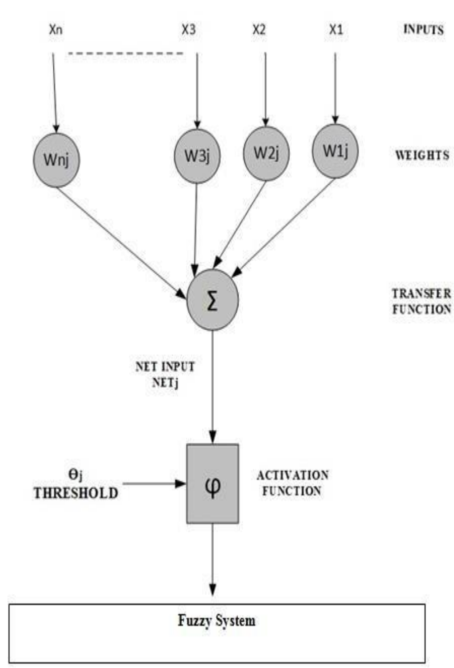

Fig.8 Block Diagram of Neuro-Fuzzy Expert Systems

3. PREVIOUS WORK

Different approaches have been devised so far for accurate time series prediction for stock process. The previous work is summarized in table1 along with their salient points.

Table.1 Summary of Previous Work

Title	Authors	Approach Used	Performance Metrics
Predicting the Direction of US Stock Prices Using Effective Transfer Entropy and Machine Learning Techniques **IEEE 2020**	S Kim, S Ku, W Chang, JW Song	Effective transfer entropy (ETE) used in conjugation with existing ML algorithms such as LR, MLP and LSTM.	Mean Absolute Percentage Error (MAPE)
Augmented Textual Features-Based Stock Market Prediction **IEEE 2019**	S Bouktif, A Fiaz, M Awad, Amir Mosavi	Opinion Mining and Sentiment Analysis was used along with historical stock prices for market prediction.	Mean Square Error, Accuracy.
Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong", Information Processing & Management **Elsevier 2018**	X Li, P Wu, W Wang	Textual normalization and opinion mining techniques were incorporated as features to gauge the sentiments of the common public regarding reputations of the firms since previous prices .	Precision, Accuracy.
Deep Learning with long short term memory networks for financial market predictions **Elsevier 2017**	Thomas Fischer, Christopher Krauss	The authors have proposed the use of long short term memory (LSTM) neural networks for stock market prediction	Root Mean Square Error (RMSE)
Financial news predicts stock market volatility better than close price", **The Journal of Finance and Data Science 2016**	Adam Atkins, Mahesan Niranjan, Enrico Gerding	Authors incorporate the Financial News data for stock market forecasting. The Bayesian approach is used.	Accuracy.
Stock market trend prediction using dynamical Bayesian factor graph **Elsevier 2015**	Lili Wang, Zitian Wang, Shuai Zhao, Shaohua Tan	A dynamical Bayesian factor graph used for prediction of Chinese Stock Exchange.	Accuracy
Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty **Elsevier 2015**	Aditya Bhardwaj, Yogendra Narayan, Vanraj, Pawan, Maitreyee Dutta	The authors incorporated sentiment analysis data from social media as a feature and found increase in accuracy of prediction	Mean Absolute Percentage Error, Accuracy.
Stock market prediction using artificial neural networks **Citeseer 2014**	B Chauhan, U Bidave, A Gangathade, S Kale	The authors have proposed the knowledge discovery process (KDD) technique and discovery mining along with neural networks for stock market prediction	Accuracy

The analysis of previously existing techniques renders an insight into the different approaches that have been prevalent thus far and hence allows us to propose a particular approach.

4. LIMITATIONS OF PREVIOUSLY EXISTING SYSTEMS AND POSSIBLE SOLUTIONS

The major challenges and limitations seen so far are:

1) Several approaches ignore the local disturbances in the mined data which can render inaccuracies in prediction.

2) Employing deep learning and back propagation simultaneously has not been explored in detail.

3) Dimensional reduction and optimization techniques such as principal component analysis (PCA) has not been coupled with the training algorithms.

Thus the proposed approach can explore using pre-processing tools like the wavelet transform defined as:

The mathematical formulation for the wavelet transform is given by the scaling and shifting approach of the wavelet function.

The scaling, shifting dependence can be defined as:

$W\varphi(Sc,Sh)$=₩[**x**,**t**] (4)

Here,
x is the space variable
t is the time variable
₩ is the transform
s_c is the scaling factor
s_h is the shifting factor

Moreover probabilistic back propagation approaches such as the Bayesian Regularization (BR) algorithm can be used to enhance accuracy. The Bayesian Regularization (BR) algorithm is a modified version of the LM weight updating rule with an additional advantage of usingthe Baye's

theorem of conditional probability for a final classification [32].
The weight updating rule for the Bayesian Regularizationis given by:

$$\boldsymbol{w_{k+1} = w_k - (J_k J_k^T + \mu I)^\wedge - 1(J_k^T e_k)} \quad \textbf{(5)}$$

Here,
w_{k+1} is weight of next iteration,
w_k is weight of present iteration
J_k is the Jacobian Matrix
J_k^T is Transpose of Jacobian Matrix
e_k is error of Present Iteration
μ is step size
I is an identity matrix.

Moreover for the predictive modelling using a data set, the Baye's Rule is followed, which is given by:

$$\boldsymbol{P\frac{A}{B} = \frac{P(A).P\frac{B}{A}}{P(B)}} \quad \textbf{(7)}$$

Here,
$P^{A/B}$ is the probability of occurrence of A given B is true.
$P^{B/A}$ is the probability of occurrence of B givenA is true.
$P(B)$ is the probability of occurrence of B
$P(A)$ is the probability of occurrence of A

5. PERFORMACNE METRICS

The performance metrics typically computed in time series applications are:
Mean Square Error:
It is mathematically defined as:

$$mse = \frac{1}{n}\sum_{I=1}^{N}(X - X')^2 \qquad (8)$$

Here,
X is the predicted value and
X' is the actual value and
n is the number of samples.
Mean Absolute Percentage Error (MAPE)

It is mathematically defined as:

$$MAPE = \frac{100}{M}\sum_{t=1}^{N}\frac{E - E_t|}{E_t} \qquad (9)$$

Here,
E_t and $E_{t\sim}$ stand for the predicted and actual values respectively.

The number of predicted samples is indicated by M. It is expected that an approach comprising of a deep learning based on the aforesaid parameters would attain lesser error and higher accuracy compared to previously existing techniques.

6. CONCLUSION

It can be concluded from previous discussions that stock market prediction is a category of time series prediction with high sensitivity and dependence on external factors. Hence it is often challenging to attain high levels of accuracy in prediction. This paper presents a comprehensive survey on the prevailing techniques employed for stock market prediction. A comparative analysis of the proposed approaches has been done. Moreover, the limitations in previously existing systems have been clearly cited. Possible solutions towards improvement of existing systems has also been proposed.

REFERENCES

1. S Kim, S Ku, W Chang, JW Song, "Predicting the Direction of US Stock Prices Using Effective Transfer Entropy and Machine Learning Techniques", IEEE Access 2020, Vol-8, pp: 111660 – 111682.

2. Jithin Eapen; Doina Bein; Abhishek Verma, "Novel Deep Learning Model with CNN and Bi-Directional LSTM for Improved Stock Market Index Prediction", 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), IEEE 2019 pp. 0264-0270.

3. Y Guo, S Han, C Shen, Y Li, X Yin, Y Bai, "An adaptive SVR for high-frequency stock price forecasting", Volume-6, IEEE Access 2018, pp: 11397 – 11404.

4. T Fischer, C Krauss, "Deep learning with long short-term memory networks for financial market predictions", European Journal of Operational Research, Elsevier 2018, Volume 270, Issue 2, , pp. 654-669

5. Amin Hedayati Moghaddam, Moein Hedayati Moghaddam, Morteza Esfandyari, "Stock market index prediction using artificial neural network", Journal of Economics, Finance and Administrative Science 21 (2016) 89–93,

6. Aditya Bhardwaj, Yogendra Narayan, Vanraj, Pawan, Maitreyee Dutta, "Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty", Procedia Computer Science 70 (2015) 85 – 91

7. Mehak Usmani, Syed Hasan Adil, Kamran Raza, Syed Saad Azhar Ali, "Stock Market Prediction using Machine Learning Techniques", DOI: 10.1109/ICCOINS.2016.7783235

8. Reza Hafezi, Jamal Shahrabi, Esmaeil Hadavandi, "A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price", Applied Soft Computing 29 (2015) 196–210

9. Xiongwen Pang, Yanqiang Zhou, Pan Wang, Weiwei Lin, Victor Chang, "An innovative neural network approach for stock market prediction", Springer Science+Business Media, LLC, part of Springer Nature 2018

10. Lili Wang, Zitian Wang, Shuai Zhao, Shaohua Tan, "Stock market trend prediction using dynamical Bayesian factor graph", Expert Systems with Applications, No. of Pages 9, 24 March 2015

11. Adam Atkins, Mahesan Niranjan, Enrico Gerding, "Financial news predicts stock market volatility better than close price", The Journal of Finance and Data Science 4 (2018) 120-137

12. Aparna Nayak, M. M. Manohara Pai and Radhika M. Pai, "Prediction Models for Indian Stock Market", Procedia Computer Science 89 (2016) 441 – 449

13. Xingyu Zhou, Zhisong Pan, Guyu Hu, Siqi Tang, and Cheng Zhao, "Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets", Hindawi Mathematical Problems in Engineering Volume 2018, Article ID 4907423, 11 pages

14. Sahar Sohangir, Dingding Wang, Anna Pomeranets and Taghi M. Khoshgoftaar, "Big Data: Deep Learning for financial sentiment analysis", J Big Data (2018) 5:3

15.Mruga Gurjar, Parth Naik, Gururaj Mujumdar, Prof. Tejaswita Vaidya, "Stock Market Prediction Using ANN", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, Volume: 05 Issue: 03 | Mar-2018

16. Raphael H. Heiberger, "Predicting economic growth with stock networks", Physica A 489 (2018) 102–111

17. Yanhui Guo, Siming Han, Chuanhe Shen, Ying Li, Xijie Yin, and Yu Bai, "An Adaptive SVR for High-Frequency Stock Price Forecasting", 2169-3536 2018 IEEE, VOLUME 6, 2018

18. Xi Zhang, Siyu Qu, Jieyun Huang, Binxing Fang, and Philip Yu, "Stock Market Prediction via Multi-Source Multiple Instance Learning", 2169-3536 2018 IEEE, VOLUME 6, 2018

19. Bruno Miranda Henrique, Vinicius Amorim Sobreiro, Herbert Kimura, "Stock price prediction using support vector regression on daily and up to the minute prices", The Journal of Finance and Data Science 4 (2018) 183-201

20. Wen Fenghua, Xiao Jihong, He Zhifang, Gong Xu, "Stock Price Prediction Based on SSA and SVM", Procedia Computer Science 31 (2014) 625 – 631