

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/341482418>

A Survey on Stock Market Prediction Using Machine Learning Techniques

Chapter · May 2020

DOI: 10.1007/978-981-15-1420-3_101

CITATIONS

36

READS

23,811

3 authors, including:



Dr Subba Rao Polamuri

Bonam Venkata Chalamayya Engineering College

25 PUBLICATIONS 128 CITATIONS

[SEE PROFILE](#)



Kudipudi Srinivas

Velagapudi Ramakrishna Siddhartha Engineering College

45 PUBLICATIONS 1,067 CITATIONS

[SEE PROFILE](#)



A Survey on Stock Market Prediction Using Machine Learning Techniques

Polamuri Subba Rao^{1(✉)}, K. Srinivas², and A. Krishna Mohan³

¹ Department of CSE, KIET, Korangi, India
psr.subbu546@gmail.com

² Department of CSE, VR Siddhartha Engineering College, Vijayawada, AP, India
vrdrks@gmail.com

³ Department of CSE, UCEK, JNTUK, Kakinada, AP, India
krishna.ankala@gmail.com

Abstract. Prediction of the Stock Market is a challenging task in predicting the stock prices in the future. Due to the fluctuating nature of the stock, the stock market is too difficult to predict. Stock prices are constantly changing every day. Estimating of the stock market has a high demand for stock customers. Applying all extracted rules at any time is a major challenge to estimate the future stock price with high accuracy. The latest prediction techniques adopted for the stock market such as Artificial Neural Network, Neuro-Fuzzy System, Time Series Linear Models (TSLM), Recurrent Neural Network (RNN) and their advantages and disadvantages are studied and analyzed in this framework work. This paper is about to discuss different techniques related to the prediction of the stock market.

Keywords: Neuron-Fuzzy systems · Artificial neural network · Hidden Markov model · Data mining · Stock market prediction · TSLM and RNN

1 Introduction

For many business analysts and researchers, forecasting the stock market price is always a challenge. Stock market prices estimation is not only an interesting but also challenging area of research. Predicting the stock market with full accuracy is very difficult as external entities such as social, psychological, political and economic have a great and substantial influence on it. The main characteristic of the data associated with stock market is usually time variant and nonlinear. Prediction of stock market plays a vital role in stock business [1]. If investors lack sufficient information and knowledge then their investment can suffer the greatest loss.

Investors must predict the future stock value of companies in order to obtain high profits. Various prediction techniques have been developed to do predictions on the stock market accurately. There were two methods widely known as conventional methods at the time when there were no computational methods for risk analysis. There are many conventional methods for predicting stock prices (by analyzing past data).

Two methods that are widely used in general are namely Fundamental Analysis and Technical Analysis.

Fundamental Analysis: To determine accurate product value, reliable and accurate information on the financial report of the company, it is necessary to have competitive strength and economic conditions in which they are interested [2]. The above value of the product can be used to make an investment decision. On the basis of this idea, “if the intrinsic value is higher than the market value it holds, invest otherwise and avoid it as a bad investment”. Not only are these parameters [3, 4] other parameters such as book value, earnings, p/e ratio, ROI etc. should be carefully analyzed to obtain an estimate of future business conditions. For the long-term predictions, Fundamental analysis is useful and the advantages are due to their systematic approach and their ability to predict changes [5].

Technical Analysis: “The idea behind technical analysis is that investors’ constantly changing attributes in response to different forces/factors make stock prices trends/movements”. Different technical factors of quantitative parameters can be used for analysis, such as trend indicators, lowest and highest daily values, indices, daily ups and downs, stock volume, etc. It is possible to extract rules from the data and the investors make future decisions based on these rules.

Different analysts may derive from the same charts different rules [5]. For both short and long term analysis, technical analysis is used. Technical analysis data is preferable over fundamental analysis data as input to system.

2 Chronicle of Stock Market

2.1 Stock Market Ground Work

The stock market brings together investors and buyers to sell and buy the shares in companies at an agreed price. Prices are determined by demand and supply. The primary market deals directly with the company’s new securities issues [6].

A share is a company’s issued document entitling its holder to be one of the company’s owners. One can get a dividend by owning a share which in turn get capital gain by selling the shares. Stock Exchanges act as a clearing house for each transaction which guarantees the payment of the security to the seller. The smooth operation of all these activities facilitates the expansion of businesses, economic growth, employment and the production of goods and services. It must be listed there to be able to trade a security on a certain stock exchange. The listing requirements are a set of conditions imposed on companies wishing to be listed by a given stock exchange. Stockbrokers are licensed agents who have access to the stock market directly to trade shares and charge the service fee. Traders buy and sell financial instruments like stocks, bonds, and derivatives. Traders can be either professional from financial institutions or a corporation, or individual investors [4]. Stock market basically serves as (1) Primary market and (2) Secondary market.

2.2 Importance of Stock Market

Indian stock market stood at third rank in the world. The Stock is essentially a share in a company's ownership. Stocks are partial ownership of businesses instead of stock tickers piece of paper, which can be traded in the stock market [4]. If company ownership is divided into 100 parts, the investor purchase one part which is equal to one share then we can own 1 percent of that company. Stock exchange uses an automated matching system driven by order. Stock prices are defined as any time how many buyers and sellers available for the same stock in the market. If the number of buyers is more than sellers then stock price becomes high and if the number of sellers higher than buyers then stock price becomes low.

The best buy and sell order are looked into a counterparty angle. The best buy order is which has the highest price and best sell order is which has the lowest price [7]. With this logic system can match the orders and executes the traders' system. SEBI (Security and Exchange Board of India) regulates the stock market. In stock markets customers preferences and requirements are different. The estimated world stock market was at \$36.6 trillion in early October 2008 [6]. The total world market for derivatives was estimated at approximately \$791 trillion in face value or nominal value, 11 times the size of the world economy [8].

3 Literature Survey

There were two important indicators in the literature for stock price forecasting. They are fundamental and technical analysis. Both were used to analyze the stock market [8, 9].

3.1 Prediction Techniques

Presented the recent methods for the prediction of stock market and give a comparative analysis [10] of all these Techniques.

Major prediction techniques such as data mining, machine learning and deep learning techniques used to estimate the future stock prices based on these techniques and discussed their advantages and disadvantages. They are,

- 3.1.1 Holt-Winters
- 3.1.2 Artificial Neural Network
- 3.1.3 Hidden Markov Model
- 3.1.4 ARIMA Model
- 3.1.5 Time Series Linear Model
- 3.1.6 Recurrent Neural Networks.

Holt-Winters, Artificial Neural Network, Hidden Markov Model are Machine Learning Techniques, ARIMA Model is Time series technique and Time series Linear Model and Recurrent Neural Networks are Deep Learning Techniques.

3.1.1 Holt-Winters

Holt-Winters is the appropriate or correct mode when the time series has trend and seasonal factors. The series was divided into three components or parts that are trend, basis and seasonality. Holt-Winters find three trend, level, and seasonal smoothening parameters. It has two variants: Additive Holt Winters Smoothening model and Multiplicative Holt-Winters model. The former is used for prediction and the latter is preferred if there are no constant seasonal variations in the series. It is mainly popular for its accuracy and in the field of prediction it has outperformed many other models. In short—term forecasts of economic development trends, Holt-Winters exponential smoothening method with the trend and seasonal fluctuations is usually used.

After removing the seasonal trends from the data, the following function is taken as an input and in return, the Holt-Winters makes the pre-calculations necessary for the purpose of forecasting. All parameters required for the forecasting purpose are automatically initialized based on the function data.

```
HWStock1_ng = HoltWinters(ds,gamma = FALSE)
predHW = predict(HWStock1_ng,n.ahead = 9)
```

3.1.2 Artificial Neural Network

An artificial neural network (ANN) is a technique inspired from biological nervous system, such as the human brain [5, 10]. It has a great ability to predict from large databases [11]. On the basis of the back—propagation algorithm, ANN is generally used to forecast the stock market. In the back—propagation algorithm, a neural network of multilayer perceptron (MLP) is used. It consists of an input layer with a set of sensor nodes as input nodes, one or more hidden layers of computation nodes and computation nodes of the output layer. These networks often use raw data and data derived from the previously discussed technical and fundamental analysis [11, 12].

A Multilayer Feed forward Neural Network is a neural network with an input layer, one or more hidden layers and an output layer. Inputs correspond to each training sample measured attributes. Inputs are passed to input layer simultaneously. The weighted outputs of these units are fed to the next layer of units that make up the hidden layer simultaneously. The weighted outputs of the hidden layers act as an input to another hidden layer, etc. The hidden layers number is an arbitrary design problem. The weighted output of the last the hidden layer acts as inputs to the output layer, which predicts the networks for certain samples.

Important parameters of NN are learning rate, momentum and epoch (Fig. 1).

Back propagation is a neural network learning algorithm [13]. The back propagation network learns by processing the sample set repeatedly and comparing the network prediction with the actual output. If the residual value exceeds the threshold value, the weight of the connections is modified to reduce the MSE between the forecast value and the original value. The weights are changed from the output layer to the first hidden layer in the opposite direction. Since the changes in the weights of the connections are made in the reverse direction, the name given to the algorithm is Back propagation [14]. Use the back propagation algorithm to perform the calculations and compare the predicted output and target output. The predicted value is not closer to the actual value and the weights are modified.

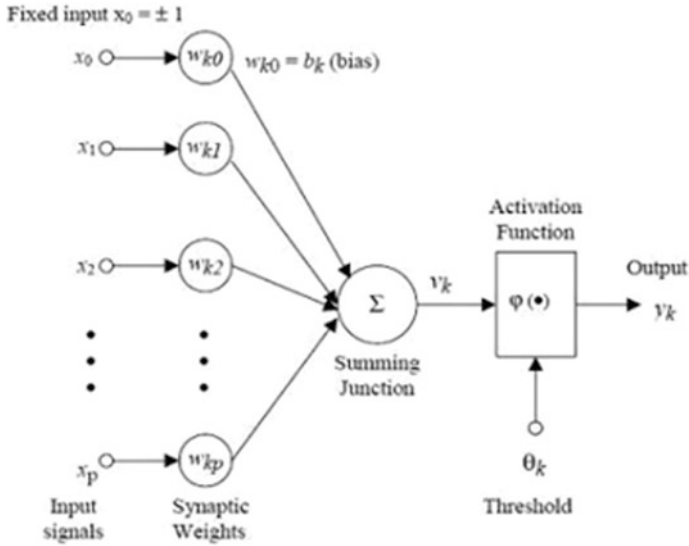


Fig. 1. Graphical representation of artificial neuron [6]

Advantages

ANN is one of the successfully developed and widely used methods for solving many prediction problems in various types of applications [9–11, 13–15].

ANNs has been used to solve various problems in financial time series forecasting and can predict the price with approximately 90% accuracy.

Disadvantages

- Neural Network is suffering from the Blackbox problem; it does not reveal the each variable's significance weight.
- The problem of overtraining is another major problem with Neural Networks [9]. The system may lose the ability to generalize if Neural Networks fits the data too well.
- Overtraining is a major problem. It usually happens for two main reasons such as Neural Networks have too many nodes or too long training periods.

3.1.3 Hidden Markov Model

In speech recognition Hidden Markov Model was first invented [16] but widely used to predict stock market related data. The stock market trend analysis is based on the Hidden Markov model, taking into account the one-day difference in close value for a given time line. The hidden sequence of states and their corresponding probability values are found for a particular observation sequence. The p probability value gives

the stock price trend percentage. In the event of uncertainty, decision-makers make decisions.

HMM is a stochastic model assumed to be a Markov Process with hidden states. It has more accuracy when compared to other models. The parameters of the HMM are indicated by A, B and p are found out.

Advantages

Hidden Markov Model gives better optimization.

Disadvantages

Hidden Markov Model's main problems are Evaluation, decoding and learning.

3.1.4 ARIMA Model

This ARIMA model was introduced by Box and Jenkins in 1970. The Box—Jenkins methodology is also referred to as a set of activities to identify, estimate and diagnose ARIMA models with time series data. The model is the most important financial forecasting method [2, 5, 16]. Models from ARIMA have been shown to be effective in generating short-term forecasts [12]. The future value of a variable in the ARIMA model is a linear combination of past values and past errors.

Advantages

- Robust and efficient forecasting of financial time series.
- Has a relatively small standard error of regression.

Disadvantages

- This model is suitable only for short term predictions.
- ARIMA models provide investors with a short-term forecast that could help to make investment decisions.

3.1.5 Time Series Linear Model

One of the stochastic ways to implement a predictive model is the linear time series model (TSLM). In a linear time series model, an ideal linear model is primarily created and data is then incorporated into it so that the linear model reflects the properties of the actual data. The main advantage of this linear model of the Time series is that the actual data are incorporated into the ideal linear model. We can include both traditional trends and seasonal data trends.

The function can be used to create the ideal linear model is in R programming is `tslm()` and incorporates `StlStock` data that have removed seasonal trends. The value `h` indicates the number of predicted or to be predicted months. The `tslm()` function performs all pre-calculations required for the prediction used as an input for the prediction function.

3.1.6 Recurrent Neural Network

Recurrent neural networks (RNN) [17] use back propagation to learn, but their nodes have a feedback mechanism. Because of this, RNN models can predict a stock price based on recent history and is recurrent [16].

```
model <- trainr(Y = trainy, X = trainx, learningrate = 0.05, hidden_dim = 5,
numepochs = 2600)
pred <- predictr(model, testx).
```

Advantage

Previous time points to input layer contains inputs.

Disadvantage

An RNN makes it possible to feed those words in through a much smaller set of input nodes.

4 Comparison of Prediction Techniques

See Table 1.

Table 1. Comparative analysis

S. no.	Techniques	Advantages	Disadvantages	Parameter used
1	Artificial neural network	Better performance compared to regression. Lower prediction error	Prediction gets worse with increased noise variation	Stock closing price
2	Support vector machine for stock prediction	Does not lose much accuracy when applied to a sample from outside the training sample	Exaggerate to minor fluctuations in the training data which decrease the predictive ability	Consumer investment, net revenue, net income, price per earnings ratio of stock, consumer spending,

(continued)

Table 1. (continued)

S. no.	Techniques	Advantages	Disadvantages	Parameter used
				diluted earnings per share, unemployment rate
3	Hidden Markov model	Used for optimization purpose	Evaluation, decoding and learning	Technical indicators
4	ARIMA	Robust and efficient	It is suitable for short term predictions only	Open, high, low, close prices and moving average
5	Time series linear model	Integrate the actual data to the ideal linear model	Traditional and the seasonal trends present in the data	Data and number of months
6	RNN	Previous time points to input layer contains inputs	It possible to feed those words in through a much smaller set of input nodes	Input hidden and output layers

5 Conclusion

This paper provides a review and comparative analysis of different stock market prediction parameter techniques. These techniques are used to evaluate stock market performance and trends. The stock market forecasting system is to increase accuracy. In this study to analyze a novel approach to improve the prediction of the results of stock, it means we will combine two or more methods to construct a novel approach method.

References

1. Naeini MP, Taremin H, Hashemi HB (2010) Stock market value prediction using neural networks. IEEE
2. Khan ZH (2011) Price prediction of share market using artificial neural network. Int J Comput Appl (IJCA)
3. Das SP, Padhy S (2014) Support vector machines for prediction of futures prices in Indian stock market. Int J Comput Appl (IJCA)
4. Chavan PS (2013) Parameters for stock market prediction. Int J Comput Technol Appl (IJCTA)
5. Sheta AF (2015) A comparison between regression, artificial neural networks and support vector machines for predicting stock market index. Int J Adv Res Artif Intell (IJARAI)
6. Preethi G, Santhi B (2012) Stock market forecasting techniques: a survey. J Theor Appl Inf Technol (JTAIT)

7. Tsai C-F, Wang S-P (2009) Stock price forecasting by hybrid machine learning techniques. In: International multiconference of engineers and computer scientists (IMECS)
8. Patel MB, Yalamalle SR (2014) Stock price prediction using artificial neural network. *Int J Innov Res Sci Eng Technol (IJIRSET)*
9. Shah M, Prabhu N, Rao J (2014) Performance analysis of neural network algorithms on stock market forecasting. *Int J Eng Comput Sci (IJECS)* 3(9). ISSN 2319-7242
10. Bharne PK, Prabhune SS (2018) Survey on combined swarm intelligences an ANN for optimized daily stock market price. In: International conference on soft computing and its engineering applications (ICSCIEA)
11. Sharaff A, Choudhary M (2018) Comparative analysis of various stock prediction techniques. In: Proceedings of the 2nd international conference on trends in electronics and informatics (ICOEI 2018) IEEE Conference
12. Samarawickrama AJP, Fernando TGI (2017) A recurrent neural network approach in predicting daily stock prices, An application to the Sri Lankan stock market. IEEE
13. Mansing GR (2014) Indian stock market prediction using neural network technique. *Int J Adv Res Comput Eng Technol (IJARCET)*
14. Sharma M (2014) Survey on stock market prediction and performance analysis. *IJIRSET*
15. Hegazy O, Soliman OS, Salam MA (2013) A machine learning model for stock market prediction. *Int J Comput Sci Telecommun (IJCST)* 4(12)
16. Ponnam LT, Srinivasa Rao V, Srinivas K (2016) A comparative study on techniques used for prediction of stock market. In: International conference on automatic control and dynamic optimization techniques (ICACDOT). International Institute of Information Technology (I²IT), Pune
17. Iqbal Z, Ilyas R, Shahzad W, Mahmood Z, Anjum J (2013) Efficient machine learning techniques for stock market prediction. *J Eng Res Appl (JERA)* 3(6). ISSN 2248-9622