

Journal of Artificial Intelligence

ISSN 1994-5450





∂ OPEN ACCESS

Journal of Artificial Intelligence

ISSN 1994-5450 DOI: 10.3923/jai.2018.48.54



Review Article Quantitative Analysis of Stock Market Prediction for Accurate Investment Decisions in Future

Surbhi Sharma and Baijnath Kaushik

School of Computer Science and Engineering, Shri Mata Vaishno Devi University, 182320 Katra, J and K, India

Abstract

Stock market is considered the primary indicator of a country's economic strength and development. Stock Market prices are volatile in nature and are affected by factors like inflation, economic growth, etc. Prices of a share market depend heavily on demand and supply. High demanded stocks will increase in price whereas heavily sold stocks will decrease in price. Fluctuating stock prices affects the investor's belief and thus there is a need to predict the future stock value. The objective of this review is to predict the stock market prices in order to make more informed and accurate investment decisions. Recent trends in stock market prediction are surveyed. Different types of machine learning classifiers and their respective variants. Various approaches and the results of past years are compared based on methodologies, datasets and efficiency and then it is represented in the form of a Graph. The survey describes different theories and conventional approaches to stock market prediction. Along with it, it discusses recent machine learning techniques along with pros and cons of each technique for effectively predicting the future stock prices followed by various researchers.

Key words: Economic growth, investment decisions, future stock value, stock market, neural networks, classifiers, shareholders, machine learning

Citation: Surbhi Sharma and Baijnath Kaushik, 2018. Quantitative analysis of stock market prediction for accurate investment decisions in future. J. Artif. Intel., 11: 48-54.

Corresponding Author: Baijnath Kaushik, School of Computer Science and Engineering, Shri Mata Vaishno Devi University, 182320, Katra, J and K, India Tel: +91-9654482709

Copyright: © 2018 Surbhi Sharma and Baijnath Kaushik. This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

A Stock market is an open market for the exchanging of organization stock at a concurred cost. It involves the trading between two investors so it is also known as Secondary Market. One of the important components of a stock market is stock exchange. A stock exchange is an organization which offers trading facilities for traders and stock brokers to trade stocks. Initially, the primary market is used for offering stocks and shares to investors and then the secondary market is used for subsequent trading¹. Here, stock exchange or broker acts as an intermediate between two parties. Stock exchange has multiple roles in economy i.e., rising capital for business, Profit sharing, investment opportunities for small investors etc. Shares and stocks are the basics of a stock market. A share or stock is a document issued by the organization which qualifies its holder for being one of the proprietors of the organization. On the other hand, a stock is a collection of shares. Common Stock and preferred stock are the two categories of a stock. Common Stock represents the possession of an organization and can claim for profits, shareholders of this type have voting rights to elect the board members. Preferred stock also represents the ownership of a company but they don't have any voting rights². Stock market prediction has been one of the challenges for researchers and financial investors stock trading is one of the problems facing by financial analysts as they are unaware of stock market behavior and they don't know which stocks to purchase and offer in order so as to acquire benefits. If the future behavior of stock prices is anticipated, they can act instantly in order to gain profits.

Therefore, the objective of this study is to predict the future stock market prices in comparison to the existing methodologies such as regression or continuous learning and by modifying them with the current methodologies efficiently by analyzing the recent trends of various researchers.

THEORIES OF STOCK MARKET PREDICTION

Various theories are available for predicting the stock market prices³. There are two important theories of stock market prediction. One is Efficient Market Hypothesis (EMH) and another one is Random Walk Theory.

Efficient Market Hypothesis (EMH): It expresses that share prices mirror all the accessible data about resources. So it is not possible to outperform the stock market. Efficient market Hypothesis exists in three forms⁴:

- Weak EMH: Only the past data is considered
- Semi-Strong EMH: All public information is utilized
- Strong EMH: Publicly and privately available information
 is used

Random walk theory: Random walk theory assumes that it is impossible to predict stock prices as stock prices don't depend on past stock. It also considers that stock price has great fluctuations so it is infeasible to predict future stock prices.

Approaches to stock market prediction: Stock market prediction have two conventional approaches^{1,2,5} (Table 1):

- Technical analysis
- Fundamental analysis

MACHINE LEARNING ALGORITHMS

Regression analysis and Hidden Markov Model: Regression Analysis is one of the non-linear methods used for stock market prediction. Regression Analysis is based on analyzing the market variables, the regression equation is set among the variables and afterward, this equation is utilized as the

Table 1: Approaches to stock market p	rediction

rable in approaches e	bioenmanerprealeann	
Parameters	Technical analysis	Fundamental analysis
Description	Future stock prices are predicted by observing the price movements of a security	Future stock prices are predicted by analyzing the economic and financial factors that affect the business
Tools of the trade	 Analysis starts with stock charts Stock prices already include all relevant information so stock charts are used rather than analyzing company's financial statements 	 Analysis starts with company's financial statements i.e., income statement, balance sheet and cash flow statement
Time horizon	Short-term approachStock charts can be delimited in weeks, days or minutes	Long-term approachFrequently looks at data over different quarters or years
Trading vs. investing	Identify short-to medium-term trades	Try to make long-term investments

predictive model to foresee the adjustments in the quantity of variables and to predict the dependent variable relationships during the forecast period.

Hidden Markov Model is also one of the methods used for predicting the stock prices. Hidden Markov Model analyzes the hidden state variables to predict the future output and state variables⁶.

Artificial neural networks: Artificial neural networks are widely used in stock market prediction. Human neurons are the basic functional unit of artificial neural networks. Neural networks can tackle an issue without an earlier learning of the connection amongst input and output, so these are also called as self-adjusting methods^{7,8}. Special function called as the activation function is used to map the input variables with output variables. In real time, neural networks have a capacity to change its network parameters (synaptic weights)⁹ neural networks are data-driven models and for real-world prediction problems like stock prediction etc¹⁰. Data-driven models are considered to be beneficial.

Naïve bayesian classifier: Machine learning is a fast-growing discipline. Machine learning is capable of integrating and acquiring the knowledge automatically. Naïve bayesian classifier falls under supervised learning method. Supervised learning method is a form of machine learning in which supervision in learning comes from labeled examples in the training dataset. Supervised learning is also called as 'Learning with the help of a teacher' because here class labels are already defined¹¹. Naïve bayesian classifier is one of the common techniques for data classification. Classification is a two-step process, one is learning step, the training set is analyzed for model construction and in classification step, class labels are predicted for the given data based on classification model.

Bayesian classifiers are statistical classifiers. Naïve bayes classifiers are based on the concept of class conditional independence i.e., impact of an attribute value on a given class is independent of the different attributes. These depend on Bayes' theorem. Bayes theorem uses the concept of posterior probability and prior probability¹².

Decision tree classifier and random forest: In decision trees, class label is represented by terminal nodes, internal nodes represent the test on an attribute and the outcome of the test is represented by branches of a tree. With the help of decision trees, classification can be easily performed. To predict the class label for a given tuple, the attribute values of the tuple

are tested against the decision tree. A path is then traced from root to a leaf node to predict the class for the given tuple. Attribute selection measures are used in decision tree classifier to choose the attribute that best partitions the tuples into particular classes. Most popular attribute selection measures used in decision tree classifier are-Information Gain, Gain Ratio and Gini Index. Decision tree classifiers have gained a lot of popularity because it can handle multidimensional data and it doesn't require any domain knowledge. In general, decision tree classifiers have good exactness¹³.

Random forests are an ensemble learning technique used for classification. Random forest is a collection of decision trees. It randomly selects the observations and specific features to build multiple decision trees and then results are calculated. In classification problems, the importance of variables is ranked by using random forests¹⁴. Few of the upsides of random forests is that there is no requirement for pruning of trees and these are not sensitive to outliers in training data. Accuracy and importance of variables are also generated automatically¹⁵.

Support Vector Machine (SVM): Support vector machines are considered to be most suitable for time series prediction. It can be used both for classification and regression task. The SVM is based on the structural risk minimization principle. This principle prevents the over-fitting problem by incorporating the concept of capacity control. Mathematical programming and Kernel Functions are the two key elements in the implementation of SVM. The SVM comes under supervised learning. Advantages of SVM is that it scales well to high dimensional data. It also reduces the computational cost because the constructed model has dependence only on support vectors. The SVM is considered as a powerful predictive tool for stock market predictions in the financial market¹⁶.

QUANTITATIVE ANALYSIS OF STOCK MARKET

Study of existing literature reviews on the basis of methodologies used for predicting stock market prices, the efficiency of existing methodologies, data sets and their efficiency are performed. The results shows that the Long Short-Term Memory (LSTM) Neural network has better results in comparison to the Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Principal Component Analysis (PCA), Word embeddings input and convolutional neural network prediction model (WB-CNN), Convolutional Neural Network (CNN) and regression methods. The result in Fig. 1 is a comparative performance efficiency of different

References	Baseline and proposed models	Datasets used	Results
Zhang <i>et al.</i> 17	SVM	China A-share, HK	China A-share HK
	PCA+SVM		ACC MCC ACC MCC
	TeSIA		55.37% 0.014 55.13% 0.08
	CMT-Z-X		57.50% 0.104 56.07% 0.092
	CMT-Z		60.63% 0.190 60.38% 0.205
	CMT		59.03% 0.162 59.36% 0.137
			60.25% 0.306 60.29% 0.252
			62.50% 0.409 61.73% 0.331
Li <i>et al.</i> ¹⁸	Long Short-Term Memory (LSTM) Neural	CS1300 index from wind database	ACC
	Network		87.86%
Gudelek <i>et al.</i> ¹⁹	Convolutional Neural Network (CNN)	Google finance	ACC
			72%
Dang and Duong ²⁰	Based on time series analysis and improved	Financial websites such as:	ACC
	text mining techniques	vietstock.vn, hsx.vn, hsn.vn	73%
Ding et al. ²¹	WB-NN	Standard and Poor's 500 stock	ACC MCC
	WB-CNN	(S and P 500) index	60.25% 0.1958
	E-CNN		61.73% 0.2147
	EB-NN		61.45% 0.2036
	EB-CNN		62.84% 0.3472
			65.08% 0.4357

J. Artif. Intel., 11 (1): 48-54, 2018

Table 2: Current papers with their evaluating parameters on respective datasets

ACC: Accuracy, MCC: Matthews correlation coefficient, SVM: Support vector machines, PCA: Principle component analysis, TeSIA: Tensor-based learning approach, CMT: Two auxiliary matrices and a tensor are factorized together, X,Z: Auxiliary matrices, WB-NN : Word embeddings input and standard neural network prediction model, WB-CNN: Word embeddings input and convolutional neural network prediction model, E-CNN: Structured events tuple input and convolutional neural network prediction model, EB-NN: Event embeddings input and standard neural network prediction model and EB-CNN : Event embeddings input and convolutional neural network prediction model



Fig. 1: Comparison of results of recent years

techniques cited among the number of research papers in the recent years. Table 2 shows the recent papers based on the use of different techniques such as SVM, KNN, PCA, WB-CNN, CNN and regression methods) along with their efficiency.

Xing *et al.*⁶ researchers have used regression analysis method and Hidden Markov Model to predict the future stock prices. They have analyzed both methods and compared their results. Although regression analysis method performs an efficient prediction, it also has great fluctuations. As stock prices change frequently, so fast and accurate prediction is must which cannot be performed with this method. Based on their experimental results, they have proved that average error

of regression analysis method is more than Hidden Markov Model. So they have concluded that Hidden Markov Model is more efficient than traditional regression analysis method in terms of accuracy because it also takes into consideration hidden variables.

Yetis, et al.²², have performed stock market prediction by utilizing artificial neural networks. They have typically focused on Multi-Layer Perceptron (MLP) networks. These are feed forward networks regularly trained with back propagation. The MLP eases the approximation of input-output map, so these are widely used for stock market prediction. Naeini et al.²³ have discussed two variants of neural networks i.e., feed forward multilayer perceptron (MLP) and an Elman recurrent network for stock market prediction. Based on the results, they have concluded that Elman network predicts the course of changes superior to multilayer perceptron but Elman recurrent network has a greater error in prediction than MLP. Usmani et al.24 have tried to predict the stock price by three variants of artificial neural networks i.e. Single Layer Perceptron Model (SLP), Multi-Layer Perceptron Model (MLP) and Radial Basis Function and by Support Vector Machine (SVM) algorithm. Single layer perceptron is the most basic arrangement which contains an input layer and an output layer. The neurons in the output layer receive the weighted sum of input neurons. Multi-layer perceptron is a feed-forward neural network with one additional layer called a hidden layer radial basis function is also fed forward network and has three layers- input, output and hidden layer. This function depends on the radial distance from a point. Based on the experimental results they have concluded that SVM performs best on training set while MLP performs best on test data set. But the prediction model works best on test data so MLP is considered to be efficient among the others for stock market prediction.

Shubhrata *et al.*¹² have performed stock market prediction using Naïve Bayes classifier. They have converted the given dataset into a frequency table and then the probabilities of events are calculated. After this, posterior probabilities of all classes are calculated using Bayes Theorem. Ultimately, a class with highest posterior probability is the outcome of prediction. Researchers have concluded that for large data sets, Naïve Bayes classifiers are assumed to be efficient as these are easy to build.

Milosevic²⁵ have predicted the stock price movement by using various algorithms like Decision Trees, Random Forests, Naïve Bayes etc. and then they have compared the accuracy of all algorithms. Based on the experimental results, it was concluded that random forests performed best as compared to other algorithms.

Kumar and Bala²⁶, Decision Trees, Random Forests and Linear Model have been used for stock market prediction. Overall study and experiments show that random forest is much better algorithm than the others due to its accuracy.

Support Vector Machine (SVM) and Back Propagation²⁷:

Techniques (BP) are used by the researchers for stock market prediction. They have compared the accuracy of both methods. Based on the experimental results, researchers have concluded that SVM performs better than BP technique as SVM provides a smaller Normalized Mean Square Error (NMSE), Mean Absolute Error (MAE) and larger directional symmetry (DS) than BPN in most cases because SVM adopts the structural risk minimization principle.

Kaushik and Banka²⁸ and Kaushik *et al.*^{29,30} proposed an approach for improving the reliability in optimal network design and fault tolerant networks. Results showed that the optimized ANN produces optimal network designs and reliability measures at reasonable computational cost.

It have summarized above work of researchers in the form of a table along with pros and cons of each technique in Table 3.

Approaches	Description	Advantages	Limitations	Comparative study in terms of accuracy
1: Regression analysis,	 Describes the relationship between a 	 Numeric forecasting method 	 Difficulty in selecting a parameter 	 Hidden Markov Model is better
Prediction method	dependent variable and one or more	 Easy to understand 	of interest	than regression analysis prediction
and	independent variables			method as it gives more accurate
Hidden Markov Model	 Takes into consideration hidden state 	 Based on implicit prediction 	 A large number of unstructured 	results in predicting future stock
	variables to predict the future state	 Avoid the influence of factor selection, 	parameters	prices
	and output variables	combination and transformation	 Inability to describe the dependencies between hidden states 	
2: Artificial neural networks	 Based on neurons 	 Self-adjusting method 	Black box	 Multi-Layer perceptron neural network
	 Ability to learn new patterns and 	 Can recognize new patterns even if not 	 High processing time for large 	is considered to be more promising in
	generalize their knowledge to	present in training set	Neural networks	stock market prediction prediction as
	predict future patterns or prices	 General function approximators 	 Requires training to operate 	compared to single layer perceptron
	 Single layer perceptron, multi-layer 	 High predictive power 		and radial basis function networks
	perceptron and radial basis function			
	are few of the variants of neural			
	networks			
3: Naïve Bayesian classifier	 Statistical classifier 	 Efficient for large datasets 	 Data scarcity 	 Efficient in predicting the future stock
	 Depends on Bayes Theorem 	 Minimum error rate 	 No curse of dimensionality 	market prices
4: Decision tree classifier	 Tree-like structure to predict the class label 	 Robust with regard to outliers 	 Not possible to predict beyond 	 Random forest is considered to be more
and	 Attribute selection measures are used to 	 Easy to interpret 	maximum and minimum limits	effective as it yields higher accuracy
	decide the best split		of the response variable	than decision tree classifier
Random forest	 An Ensemble Learning method for 	 Efficient to use for complex classification 	 Might be slow due to a large 	
	classification	tasks	number of trees	
	 Used to rank the importance of variables 	 Created model can be easily interrupted 		
5: Support Vector	 Based on structural risk minimization principle 	 High accuracy 	 Hard to interpret 	 Considered as a powerful
Machine (SVM)	Used for both classification and regression	Kernel's are available for	 Time-consuming in case non-linear data 	predictive tool for stock market of large support vectors predictions

CONCLUSION AND FUTURE SCOPE

In this study, stock market basics are discussed and then the need for predicting the future stock market prices. Few of the approaches which may be used for stock market prediction like Non-linear regression analysis, Hidden Markov Model, Artificial Neural Networks, Naïve Bayes Classifier, Decision Trees Classifier, Random Forest Method, Support Vector Machines, PCA (Principal Component Analysis), WB-CNN (Word embeddings input and convolutional neural network prediction model) and CNN (Convolutional Neural Network) are elaborated in this paper. Results of this research are beneficial in concluding that LSTM (Long Short-Term Memory) Neural network has better results in comparison to other methods.

As a future direction, this research would like to perform a comparative analysis with deep learning classifiers and extreme learning classifiers with the help of a feature reduction algorithm based on the parameters used for stock market prediction. Along with this, research would also like to study and implement economic growth model for stock market prediction and the analysis of how economic growth model will affect in stock market prediction in comparison to the linear regression model and with specialized machine learning techniques.

SIGNIFICANCE STATEMENT

In this paper, it is discovered that Stock Market Prediction is an important issue for financial investors to decide which stocks one should buy and sell. Comparative performance efficiency of different techniques based on methodologies and datasets are beneficial in modifying them with the current methodologies for efficiently predicting the stock market prices. This study will help researchers in understanding the different machine learning approaches used till now along with pros, cons and their performance efficiency. This concise information will help them to explore other possibilities. Research will be helpful in implementing the economic growth model in future for stock market prediction.

REFERENCES

- 1. Soni, S., 2011. Applications of ANNs in stock market prediction: A survey. Int. J. Comput. Sci. Eng. Technol., 2:71-83.
- Setty, D.V., T.M. Rangaswamy and K.N. Subramanya, 2010. A review on data mining applications to the performance of stock marketing. Int. J. Comput. Applic., 1: 33-43.

- Falinouss, P., 2007. Stock trend prediction using news articles: A text mining approach. Master's Thesis, Lulea University of Technology, Sweden.
- Attigeri, G.V., P.M.M. Manohara, R.M. Pai and A. Nayak, 2015. Stock market prediction: A big data approach. Proceedings of the IEEE Region 10th Conference on TENCON 2015, November 1-4, 2015, Macao, China, pp: 1-5.
- 5. Rechenthin, M.D., 2014. Machine-learning classification techniques for the analysis and prediction of high-frequency stock direction. Ph.D. Thesis, University of Iowa, USA.
- Xing, T., Y. Sun, Q. Wang and G. Yu, 2013. The analysis and prediction of stock price. Proceedings of the IEEE International Conference on Granular Computing (GrC), December 13-15, 2013, Beijing, China, pp: 368-373.
- Kaushik, B. and H. Banka, 2015. Performance evaluation of Approximated Artificial Neural Network (AANN) algorithm for reliability improvement. Applied Soft Comput., 26: 303-314.
- Kaushik, B., N. Kaur and A.K. Kohli, 2015. Improved neural approach in maximising reliability for increasing networks. Int. J. Comput. Sci. Eng., 11: 176-185.
- Majumder, M. and M.A. Hussian, 2007. Forecasting of Indian stock market index using artificial neural network. Inform. Sci., 2007: 98-105.
- Buche, A. and M.B. Chandak, 2016. Stock market prediction using text opinion mining: A survey. Int. J. Adv. Res. Comput. Sci. Software Eng., 6: 566-569.
- 11. Han, J., J. Pei and M. Kamber, 2011. Data Mining: Concepts and Techniques. 3rd Edn., Elsevier, Amsterdam, Netherlands, Pages: 702.
- Shubhrata, D.M., V.D. Kaveri, R.T. Pranit, Y.S. Bhavana and P.J. Chate, 2016. Stock market prediction and analysis using naive bayes. Int. J. Recent Innov. Trends Comput. Commun., 4: 121-124.
- Cheng, C.H. and Y.S. Chen, 2007. Fundamental analysis of stock trading systems using classification techniques. Proceedings of the International Conference on Machine Learning and Cybernetics, Volume 3, August 19-22, 2007, Hong Kong, China, pp: 1377-1382.
- 14. Prajwala, T.R., 2015. A comparative study on decision tree and random forest using R tool. Int. J. Adv. Res. Comput. Commun. Eng., 4: 196-199.
- 15. Horning, N., 2013. Introduction to decision trees and random forests. Am. Mus. Nat. Hist., 2: 1-27.
- Patil, S.S., K. Patidar and M. Jain, 2016. Stock market prediction using support vector machine. Int. J. Curr. Trends Eng. Technol., 2: 18-25.
- Zhang, X., Y. Zhang, S. Wang, Y. Yao, B. Fang and S.Y. Philip, 2018. Improving stock market prediction via heterogeneous information fusion. Knowl.-Based Syst., 143: 236-247.

- Li, J., H. Bu and J. Wu, 2017. Sentiment-aware stock market prediction: A deep learning method. Proceedings of the 2017 International Conference on Service Systems and Service Management (ICSSSM'17), June 16-18, 2017, IEEE, Dalian, China, ISBN:978-1-5090-6371-0, pp: 1-6.
- Gudelek, M.U., S.A. Boluk and A.M. Ozbayoglu, 2017. A deep learning based stock trading model with 2-D CNN trend detection. Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), November 27-December 1, 2017, Honolulu, HI, USA., pp: 1-8.
- Dang, M. and D. Duong, 2016. Improvement methods for stock market prediction using financial news articles. Proceedings of the 2016 3rd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS), September 14-16, 2016, Danang, Vietnam, pp: 125-129.
- Ding, X., Y. Zhang, T. Liu and J. Duan, 2015. Deep learning for event-driven stock prediction. Proceedings of the 24th International Joint Conference on Artificial Intelligence, July 25-31, 2015, AAAI Press, pp: 2327-2333.
- Yetis, Y., H. Kaplan and M. Jamshidi, 2014. Stock market prediction by using artificial neural network. Proceedings of the 2014 International Conference on World Automation Congress (WAC), August 3-7, 2014, IEEE, Waikoloa, Hawaii, USA., ISBN:978-1-8893-3549-0, pp: 718-722.
- 23. Naeini, M.P., H. Taremian and H.B. Hashemi, 2010. Stock market value prediction using neural networks. Proceedings of the International Conference on Computer Information Systems and Industrial Management Applications, October 8-10, 2010, Krackow, Germany, pp: 132-136.

- Usmani, M., S.H. Adil, K. Raza and S.S.A. Ali, 2016. Stock market prediction using machine learning techniques. Proceedings of the 3rd International Conference on Computer and Information Sciences (ICCOINS'16), August 15-17, 2016, IEEE, Kuala Lumpur, Malaysia, ISBN:978-1-5090-2550-3, pp: 322-327.
- 25. Milosevic, N., 2016. Equity forecast: Predicting long term stock price movement using machine learning. https://arxiv.org/ abs/1603.00751
- Kumar, P. and A. Bala, 2016. Intelligent stock data prediction using predictive data mining techniques. Proceedings of the International Conference on Inventive Computation Technologies (ICICT), August 26-27, 2016, Coimbatore, India, pp: 1-5.
- Das, S.P. and S. Padhy, 2012. Support vector machines for prediction of futures prices in Indian stock market. Int. J. Comput. Applic., 41: 22-26.
- 28. Kaushik, B. and H. Banka, 2014. Approach for improving reliability in optimal network design. Int. J. Adv. Intell. Paradigms, 6: 157-175.
- 29. Kaushik, B., N. Kaur and A.K. Kohli, 2013. Achieving maximum reliability in fault tolerant network design for variable networks. Applied Soft Comput., 13: 3211-3224.
- Kaushik, B., N. Kaur and A.K. Kohli, 2013. Improved approach for maximizing reliability in fault tolerant networks. J. Adv. Comput. Intell. Intell. Inform., 17: 27-41.