A SURVEY ON STOCK MARKET PREDICTION USING MACHINE LEARNING TECHNIQUES FOR NOISE REDUCTION.

M.Dhivya
Research Scholar, Department of Computer Science,
A.V.V.M Sri Pushpam College (Autonomous), Poondi, Thanjavur (Dt)
Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.

Dr.V.Maniraj
Research Advisor and Co-ordinator,
Department of Computer Science,
A.V.V.M Sri Pushpam College (Autonomous), Poondi, Thanjavur (Dt),
Affiliated to Barathidasan University, Tiruchirappalli, Tamil Nadu, India.

ABSTRACT

Stock market prediction is a crucial and challenging task due to its nonlinear, evolutionary, complex, and dynamic nature. Research on the stock market has been an important issue for researchers in recent years. Companies invest in trading the stock market. Predicting the stock market trend accurately will minimize the risk and bring a maximum amount of profit for all the stakeholders. During the last several years, a lot of studies have been done to predict stock market trends using Traditional, Machine learning and deep learning techniques.To make predictions regression analysis is used mostly. In this paper we survey of well-known efficient regression approach to predict the stock market price from stock market data based. This survey will assist the readers & researchers in selecting algorithms that can be useful for a predicting the stock market. A survey of various algorithms and its parameters for stock market prediction is presented in this paper.

INTRODUCTION

STOCK MARKET

The Stock market plays a vital role in the country's economic growth as well as the individual economy to a large extent. Finding the right time to buy and sell the shares is dependent on predicting the trends in the stock market. The technique for most accurate prediction is to learn from past instances and design a model to do this by using traditional & machine learning algorithms[1]. The Stock market trend varies due to several factors such as political, economics, environment, society, etc. [2][3][18][19]. There are two types of stock analysis. One is a fundamental analysis, which requires study of the company's basics such as balance sheet, expenses and revenues, annual returns, company's profile, and position, etc.

The other one is a technical analysis ,which deals with studying the statistics generated by market activities such as historical data, past price, and volumes [4]. There are two essential theories used in conventional approach for stock market prediction namely Efficient Market Hypothesis (EMH) introduced by Fama in 1964 which states that stock price future is unpredictable based on the historical data [5] and Random Walk Hypothesis (RWH) which states that stock's future price is independent of its history. Tomorrow's stock price has nothing to do with today's worth, but tomorrow's information [2][26]. The objective of this paper is to compare the traditional , Machine and Deep learning algorithms and assist the researchers for further analysis.

The stock market is the place where the stocks are transferred, traded and circulated. On the one hand, the issuance of stock provides a legal and reasonable channel for capital flow, which enables a large amount of idle capital to be gathered in the stock market. Such effective accumulation of capital can improve the organic composition of enterprise capital and greatly promote the development of economy. On the other hand, the circulation of stock enables the capital to be collected effectively and the accumulation of capital is effectively promoted. Based on this, the stock market is generally regarded by scholars from all walks of life as an intuitive reflection of the economic development of a country or region in a certain period. One of the main reasons lies in the stock market trading prices can objectively reflect the stock market supply and demand relations. Moreover, the stock market is often regarded as an indicator of stock prices and quantities. However, due to the complexity, variability and uncertainty of the stock market, the stock price formation mechanism presents the characteristics of complexity and unpredictability.

STOCK MARKET ANALYSIS

Stock prices not only from the political, economic, market, technology, and investor behavior aspects such as individual factors, influenced by various factors in the interaction between the role at the same time, these will lead to changes in stock prices, the existence of the various uncertain factors lead to the complexity of the stock price changes. It is the constant changes in stock prices that provide the breeding space for speculation in the capital market and increase the risk of the stock market. Investors and analysts individuals tend to be "irrational", relying on personal experience and intuition to make decisions and determine there is a certain limitation, namely with experience and intuition to forecast the stock price is not accurate, this is

not accuracy under the guidance of related behavior caused greater risks, could lead to economic losses to investors. Therefore, how to accurately analyze, judge and predict stock prices for investors to make decisions is very critical [1-10].

get ultimate conclusions.

MACHINE LEARNING

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.Machine learning is used in internet search engines, email filters to sort out spam, websites to make personalised recommendations, banking software to detect unusual transactions, and lots of apps on our phones such as voice recognition.

DEEP LEARNING

Deep learning based on neural network has attracted extensive attention from scholars in the field of deep learning. Neural network is a complex nonlinear dynamic system, therefore, what other method can't deal with low efficiency of complex and nonlinear system can be through neural network method, this method is characterized by handling mechanism can highly parallel, topological structure performance is very flexible, operation ability is extremely powerful, nonlinear operation faster and stronger ability of self learning organization. This method has been paid attention to in the research field.

LITERATURE SURVEY

This section discusses the similar work for stock market prediction, implemented by various authors using different algorithms. There are many datasets that can be used for data related to stocks. For example, S&P 500, NASDAQ, Karachi Stock Exchange (KSE), London Stock Exchange (LSE), NewYork Stock Exchange (NSE). Cao and Tay [1] used the S&P 500 dataset. Time series forecasting is an essential area in which variables past observations are collected and analyzed for development of the model, to describe the underlying relationships, so that the model can be used to predict the future [10]. ARIMA is one of the widely used time series models due to its statistical properties, and it can represent various types of time series like Pure Autoregressive (AR), Pure Moving Average (MA) and combined AR and MA (ARMA) series [17]. Univariate model ARIMA was parametric and based on the assumption that the nature of time series is linear and stationary. ARIMA failed to capture the nonlinear patterns because it is a linear model. To overcome this

limitation , ANN was used due to its flexible nonlinear capabilities [17]. ANN's performance was found to be satisfactory because it could rely on more considerable information, technical indicators, fundamental indicators, etc.But ANN had some critical issues like overfitting , which lead to poor generalization and out of sample data. To resolve the mentioned issues the procedures such as Cross -Validation, Non-parametric Probability Density Estimation [6], Modifying Training Algorithms and Pruning parameters or hidden notes[7] were used.

To take advantage of the unique strength of both the models, Zhang proposed a hybrid model comprising of combination of both ARIMA and ANN[10]. In hybrid model, ARIMA was used to analyze the linear part of the problem, and ANN was used to model the residuals provided by ARIMA. By using a hybrid model , one could model linear and nonlinear patterns separately and then combine the forecast with improving the model's forecast and performance. Three datasets namely, Wolf's sunspot data, the Canadian lynx data, and British Pound/US dollar exchange rate data were used. To forecast the accuracy performance, MSE and MAD were selected. Based on the evaluation measures, it was concluded that the hybrid model outer performed each component model used separately. Another version of ANN, Dynamic ANN (DAN2), was used by Guresen et al.[12] using Multilayer perceptron (MLP) and Hybrid Neural Networkwhich used Generalized Auto-Regressive Conditional Heteroscedasticity (GARCH) to extract new features. The dataset used as NASDAQ, and the evaluation of the models was done by using two evaluation metrics namely, MSE and MAD. The paper concluded that the overall results showed that the ANN model MLP gave the best and reliable result in time series forecasting , whereas the hybrid model failed to improve the outcome.

Zhu et al. [11] tried to investigate whether the trading volume features could improve the neural network performance under short, medium , and long-term forecasting scale. Component- based three-layer Feed Forward Neural Network (FFNN) was employed , and trading volume under different input selection schemes with the basic models was set up as three augmented network models to test if trading volume could significantly improve the performance. Three datasets namely, NASDAQ, DJIA , and STI, were used to investigate the impact of trading volume on daily, weekly and monthly predictions. The observations of the experiment were that trading volume could not fundamentally improve the forecasting performance for the short term, in

contrast it modestly improved for the medium term, but the augmented model gave the best prediction for the long term.

III. STOCK MARKET PREDICTION

Prediction of stock prices is very challenging and complicated process because price movement just behaves like a random walk and time varying. In recent years various researchers have used intelligent methods and techniques in stock market for trading decisions. Here, we present a brief review of some of the significant researchers. A Sheta [7] has used TakagiSugeno (TS) technique to develop fuzzy models for two nonlinear processes. They were estimated software effort for a NASA software projects and the prediction of the next week S&P 500 for stockmarket. The development process of the TS fuzzy model can be achieved in two steps 1) the determination of the membership functions in the rule antecedents using the model input data; 2) the estimation of the consequence parameters. They used least-square estimation to estimate these parameters.

The results were promising. M.H. FazelZarandiet al. [8] have developed a type-2 fuzzy rule based expert system for stock price analysis. Interval type-2 fuzzy logic system permitted to model rule uncertainties and every membership value of an element was interval itself. The proposed type-2 fuzzy model applied the technical and fundamental indexes as the input variables. The model can be tested on stock price prediction of an automotive manufactory in Asia. Robert K. Lai et al. [9] have established a financial time series-forecasting model by evolving and clustering fuzzy decision tree for stocks in Taiwan Stock Exchange Corporation (TSEC). The forecasting model integrated a data clustering technique, a fuzzy decision tree (FDT), and genetic algorithms (GA) to construct a decision-making system based on historical data and technical indexes. The set of historical data can be divided into k subclusters by adopting K-means algorithm. GA was then applied to evolve the number of fuzzy terms for each input index in FDT so the forecasting accuracy of the model can be further improved. S AbdulsalamiSulaiman Olaniyi et al [11] have proposed a linear regression method of analyzing coupled behavior of stocks in the market. The method successfully predicts stock prices based on two variables.

IV. PREDICTION METHODS

REGRESSION

In statistics, regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables . More specifically, regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the conditional expectation of the dependent variable given the independent variables – that is, the average value of the dependent variable when the independent variables are fixed. Less commonly, the focus is on a quantile, or other location parameter of the conditional distribution of the dependent variable given the independent variables. In all cases, the estimation target is a function of the independent variables called the regression function. In regression analysis, it is also of interest to characterize the variation of the dependent variable around the regression function which can be described by a probability distribution. Regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning.

Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships. In restricted circumstances, regression analysis can be used to infer causal relationships between the independent and dependent variables. However this can lead to illusions or false relationships, so caution is advisable;[1] for example, correlation does not imply causation. Many techniques for carrying out regression analysis have been developed. Familiar methods such as linear regression and ordinary regression are parametric, in that the regression function is defined in terms of a finite number of unknown parameters that are estimated from the data. Nonparametric regression refers to techniques that allow the regression function to lie in a specified set of functions, which may be infinitedimensional.

The performance of regression analysis methods in practice depends on the form of the data generating process, and how it relates to the regression approach being used. Since the true form of the data-generating process is generally not known, regression analysis often depends to some extent on making assumptions about this process. These assumptions are sometimes testable if a sufficient quantity of data is available.

Regression models for prediction are often useful even when the assumptions are moderately violated, although they may not perform optimally. However, in many applications, especially with small effects or questions of causality based on observational data, regression methods can give misleading results.[2][3] In a narrower sense, regression may refer specifically to the estimation of continuous response variables, as opposed to the discrete response variables used in classification.[4] The case of a continuous output variable may be more specifically referred to as metric regression to distinguish it from related problems.

NOISE REDUCTION

A coherence-based dual-microphone noise reduction technique and showed that in anechoic (also low reverberant) rooms, where the noise field is highly coherent, it offers significant improvements over a fixed directional microphone and a well-established beamformer in terms of intelligibility and quality. We also observed that algorithm performance starts to degrade when tested in more reverberant rooms. The main reason is that the algorithm assumes that signals received by the two microphones are purely coherent. Although this assumption is valid for low reverberant environments, the coherence function captures the characteristics of diffuse noise when the (late) reverberation is substantial. In this work, we modify the modeling of the coherence function utilized in [8], in such a way so that it takes into account both the coherent and diffuse noise of the environment (i.e., a hybrid coherence model). This modification is achieved by estimating and applying the direct-to-reverberant energy ratio (DRR) into the coherence model presented in [8]. The DRR represents the ratio between the received signals corresponding to the direct path (i.e., coherent signal) and those subject to the multipath reflections (diffuseness). In this letter, we will show that by improving the basic coherence model presented in [8], the new algorithm achieves better noise reduction performance in more reverberant environments.

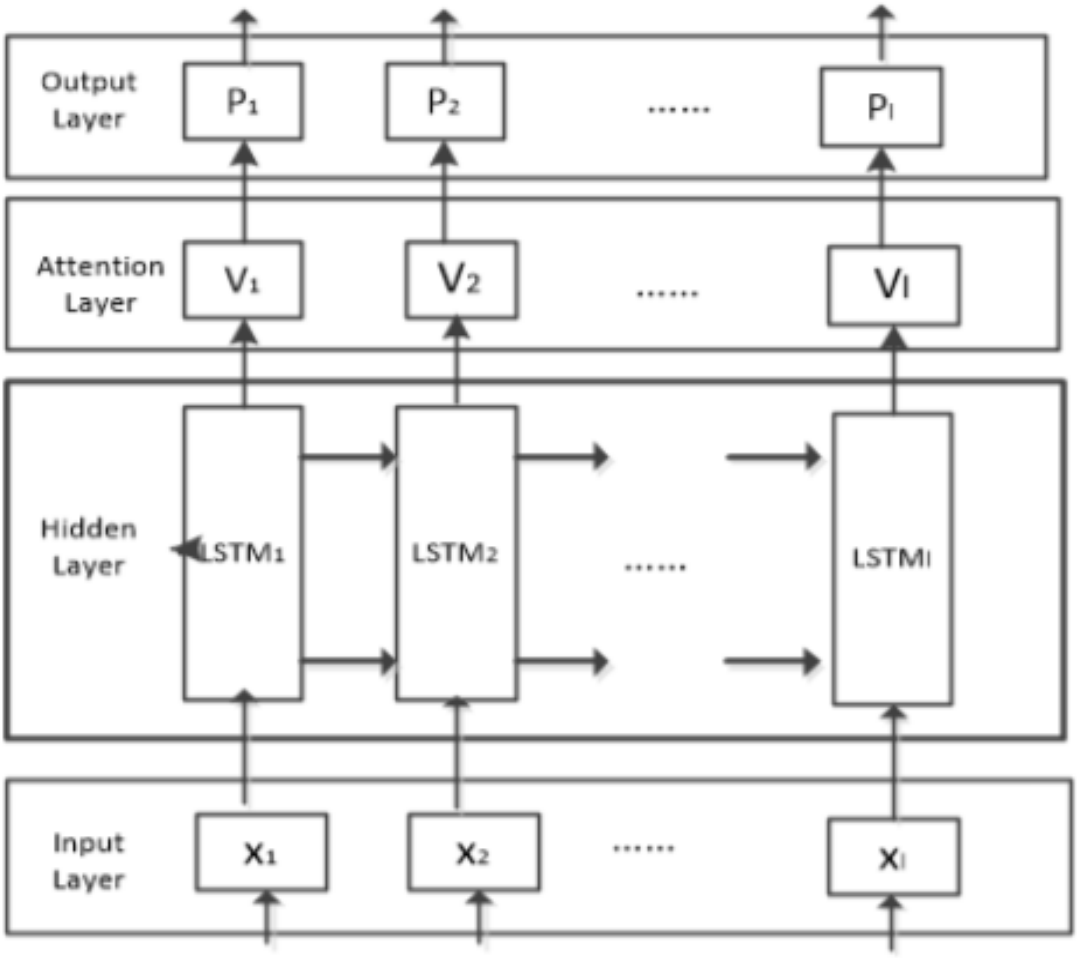

Figure.1. LSTM MODEL

Table I. Comparative Of Dataset, Algorithm, Features And Evaluation Metrics Used

S.No	DatasetUsed	Algorithmsused	Evaluation MetricsUsed	Features
1	S&P500	MLP-BPSVM	NMSEMAE DS CPCD	Days,RelativeFrequency
2	S&P500	NN-BPLS-SVM LS-SVMPSO	MSE	Days,ClosingPrice
3	The Wolf'ssunspot data theCanadian lynxdataBritish Pound/US dollarexchangeratedata	ARIMAANN Hybrid	MSEMAD	Days,Price
4	NASDAQDJIA STI	Component- basedthree-layer feedforward neuralnetwork	One-stepSignpredictionrate MSE	TradingVolume, Price
5	NASDAQ	MLP DAN2GARCH	MSEMAD	Day,Price
6	Karachi, Londonand New Yorkstockexchange	KNNSVM NaïveBayes	MAERMSE Accuracy	Date, Open,Low, High,Close,Volume,Trend,Sentiment & Futuretrendvalue
7	DJIA	RNN-LSTMCNN	MeanAccuracy	Price andDate,News
8	NSE,NYSE	MLPRNNLSTMCNN	MAPE	ClosingPrice,Days
9	BSE	PSOLSM	MAPE	ClosingValues, Days
10	IBMInc	SVM Radial BasisFunction	Log2cLog2g	Pricevolatility,StockMomentum,Indexvolatility,IndexMomentum

The evaluation metrics used for stock prediction are NMSE,MAE,DS,CP,CD, MSE and MAD. Stock prediction is made based on multiple features like closing price, days, relative frequency, trading volume, price volatility, stock momentum, index volatility, index momentum. Various algorithms, evaluation metrics , and features used by different datasets are summarized in table 1.

V.CONCLUSION

This paper aims to study the stock market prediction using multiple Traditional, Machine learning, and Deep learning algorithms. Along with the algorithms, the survey has focused on various datasets used for stock market prediction, features of these datasets selected as input parameters and the evaluation metrics used for comparing the results of predictions.In future the results of multiple regression approach could be improved using more number of variables. using LSTM neural network model will be implemented, such as the time lag of prediction, but with attention layer, it can predict stock prices. Its main principle is to discover the role of time series through analyzing the historical information of the stock market, and to deeply explore its internal rules through the selective memory advanced deep learning function of LSTM neural network model, so as to achieve the prediction of stock price trend.

VI.REFERENCES

1) Lijuan Cao and Francis E.H. Tay, “Financial Forecasting Using Support Vector Machines”, 2001 Springer-Verlag London Limited.
2) Jinqi Tang, Xiong Chen, “Stock Market Prediction Based on Historic Prices and News Titles”, ACM ISBN 978-1-4503-6432- 4/18/05.
3) Billah, M., and Waheed, S. 2017. “Stock market prediction using an improved training algorithm of neural network”, International Conference on Electrical, Computer and Telecommunication Engineering. pp. 1-4. IEEE.
4) Nischal Puri, Avinash Agarwal, Prakash Prasad, “A Survey on Machine Learning Approach for Stock Market Prediction”, Helix Vol. 8(5): 3705- 3709.
5) Chang Sim Vui, Gan Kim Soon, Chin Kim On, and Rayner Alfred, “A Review of Stock Market Prediction with Artificial Neural Network (ANN)”, 2013 IEEE International Conference on Control System, Computing and Engineering, 29 Nov. - 1 Dec. 2013, Penang, Malaysia..
6) Pan ZH, Wang XD. “Wavelet-based density estimator model for forecasting”, J Computational Intelligence in Finance 1998; 6: 6–13.

7) Dorsey R, Sexton R. “The use of parsimonious neural networks for forecasting financial time series”, J Computational Intelligence in Finance 1998; 6: 24–30.

8) Vapnik VN. “The Nature of Statistical Learning Theory”, New York, Springer-Verlag, 1995.

9) Osman Hegazy , Omar S. Soliman , and Mustafa Abdul Salam, “A Machine Learning Model for Stock Market Prediction”, International Journal of Computer Science and Telecommunications volume 4, Issue 12, December 2013.

10) G. Peter Zhang, “Time series forecasting using a hybrid ARIMA and neural network model”,2002 Elsevier Science B.V. All rights reserved.

11) Xiaotian Zhu, Hong Wang , Li Xu , Huaizu Li, “Predicting stock index increments by neural networks: The role of trading volume under different horizons”, 2007 Elsevier Ltd. All rights reserved.

12) Erkam Guresen, Gulgun Kayakutlu ,Tugrul U. Daim , “Using artificial neural network models in stock market index prediction”, 2011 Elsevier Ltd. All rights reserved.

13) Khan, W. Ghazanfar, M. A. Asam, M, Iqbal, A. ,Ahmad, S. ,Javed Ali Khan, “Predicting trend in stock market exchange using machine learning classifiers”, Sci.Int.(Lahore),28(2),1363- 1367,2016.

14) Hiransha M, Gopalakrishnan E.Ab, Vijay Krishna Menon, Soman K.P, “NSE Stock Market Prediction Using Deep-Learning Models”, 1877- 0509 © 2018 The Authors. Published by Elsevier Ltd. and Data Science (ICCIDS 2018).

15) Asif Perwej, K. P. Yadav, Vishal Sood, Yusuf Perwej, “An Evolutionary Approach to Bombay Stock Exchange Prediction with Deep Learning Technique”, IOSR Journal of Business and Management (IOSR-JBM)

16) V Kranthi Sai Reddy, “ Stock Market Prediction Using Machine Learning”, International Research Journal of Engineering and Technology, Volume: 05 Issue: 10 | Oct 2018

17) G.E.P. Box, G. Jenkins, “Time Series Analysis, Forecasting and Control, Holden-Day”, San Francisco, CA, 1970.

18) Chen. K. Zhou. Y., & Dai, F. 2015. A LSTM-based method for stock returns prediction: A case study of China stock market. IEEE International Conference on Big Data. 2823-2824. IEEE.

19) Mahalakshmi, G., Sridevi, S,. & Rajaram, S. 2016. A Survey on forecasting of time series data. International Conference on Computing Technologies and Intelligent Data Engineering. 1-8. IEEE.

20) S. SElvin, R. Vinayakumar, E.A. Gopalakrishnan, V.K. Menon and K.P. Soman. (2017) "Stock Price Prediction using LSTM, RNN and CNN- sliding window model." International Conference on Advances in Computing, Communications and Informatics; 1643-1647.

21) Rather A. M., Agarwal A., and Sastry V. N. (2015). :Recurrent neural network and a hybrid model for prediction of stock returns." Expert System with Applications 42(6): 3234-3241.

22) Zhang G., Patuwo B. E., and Hu M/ Y. (1998). "Forecasting with artificial neral networks: The State of Art." International Journal of Forecasting 14(1): 35-62.

23) Mittermayer, Marc-Andre. "Forecasting intraday stock price trends with text mining techniques." System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on IEEE, 2004.

24) Schumaker, Robert P., H. Chen. "Textual analysis of stock market prediction using breaking financial news: The AZFin Txt System." ACM Transactions on Information Systems (TOIS) Volume 27, Issue 2, pp:12, 2018.

25) Shen, Shunrong, H. Jiang, T. Zhang. "Stock Market Forecasting using machine learning algorithms." , 2018.

26) O, Bernal. S, Fok. R, Pidaparthi. 2018. Financial Market Time Series Prediction with Recurrent Neural Networks. Citeseer.

27) Amin Hedayati Moghaddam, Moein Hedayati Moghaddam, Morteza Esfandyari ," Stock Market Index Prediction Using Artificial Neural Network",2077-1886© 2016 Universidad ESAN, Elsevier.

28) Omer Berat Sezera,c, Murat Ozbayoglua1, Erdogan Dogdub ," A Deep Neural-Network Based Stock Trading System Based on Evolutionary Optimized Technical Analysis Parameters ", Procedia Computer Science 114 (2017) 473–480.

29) Tian Xia, Qibo Sun, Ao Zhou, Shangguang Wang, Shilong Xiong, Siyi Gao, Jinglin Li, Quan Yuan, «Improving the Performance of Stock Trend Prediction by Applying GA to Feature Selection", 2018 IEEE 8th International Symposium on Cloud and Service Computing (SC2).

30) Xi Zhang 1, (Member, Ieee), Siyu Qu1, Jieyun Huang 1, Binxing Fang1, And Philip Yu2, “Stock Market Prediction via MultiSource Multiple Instance Learning”, 2169-35362018 IEEE. Translations