International Journal of Research in Engineering, Science and Management
Volume-2, Issue-2, February-2019
www.ijresm.com | ISSN (Online): 2581-5792

Forecasting Stock Market Future Movement Direction: Supervised Machine Learning Algorithm

S. Yasmeen[1], M. Meera[2], K. Gowthami[3], D. Lakshmi[4]
[1,2,3]*Student, Department of CSE, Panimalar Institute of Technology, Chennai, India*
[4]*Assistant Professor, Department of CSE, Panimalar Institute of Technology, Chennai, India*

***Abstract*: Time series forecasting has been widely used to determine the future trends of stock and the analysis and modeling of finance time series importantly guide investor's decisions and trades. In a financially volatile market, it is important to have a very precise direction of a future stock. Because of the financial crisis and gaining profits, it is necessary to have a secure prediction of the value of the stocks. Predicting a non-linear signal requires an advanced algorithms of machine learning. So, this system will analyze the current rate of the products and stocks. System will recommend the high profitable product when the stock rate is increase and also it will recommend which product can put on sale to get a profit. This project proposes a new prediction algorithm that exploits the temporal correlation among global stock markets and various financial products to predict the next day stock trend with the aid of SVM along with Alert System. This algorithm is also applied with different regression algorithms to treat the actual increment in markets. The proposed system is validated by comparing its performance with those of empirical methods and previous works through hypothesis test via real world engineering cases. The obtained results indicate that the proposed scheme will reduce false positives to great extent and is practical enough to be implemented in real time scenarios. Finally, a simple trading model is established to study the performance of the prediction algorithm against other benchmarks.**

***Keywords*: Time series, prediction, Forecast, Machine Learning, Support Vector Machine (SVM), Investors, Non-linear signal, Stock, Investment, Alert.**

1. Introduction

Financial predictions are one of the hottest topics not only for researches but also for commercial applications. Stock market is a very complex non-linear time series system. It is assumed that the behaviors of stock market to be shown in the future could be predicted with previous information given in the history. The traditional methods used to forecast indices and prices of stock market, such as auto regressive model, moving averages model, exponential weighted moving averages model and generalized autoregressive conditional heteroscedasticity model etc., are all based on probability theory and statistical analysis with a certain of distributions assumed in advance. It is easy to see that these assumptions are unreasonable and non-realistic.

Moreover, the statistical models are more or less lack of accuracy of prediction due to the linear structure of modeling system. It relies on the detection of strong empirical regularities in the observations of the system, with the assumption that any influences on stock prices are already reflected in the price movements. Predicting such highly fluctuating and irregular data is usually subject to huge errors. So, developing more realistic models for predicting financial time series data is a great interest in financial time series. Machine learning, a well-defined algorithm in a wide range of applications, has been extensively studied for its potentials in prediction of financial stocks. Popular algorithms, including support vector machine (SVM), have been reported to be quite effective in tracing the stock market and help maximizing the profit of stock purchase while keeping the risk low. However, in many of these literatures, the selected inputs to the machine learning algorithms are mostly derived from the data within the same market. Such isolation leaves out important information carried by other entities and make the prediction result more vulnerable to local commotion. Efforts have been done to break the boundaries by incorporating external information through fresh financial news or personal internet posts such as Facebook, Twitter, Instagram, etc. This approach is known as sentimental analysis, replies on the attitudes of several key figures or successful analysts in the markets to interpolate the minds of general investors. Despite its success in some occasions, sentimental analysis may fail when some of the people are prejudge mental, or positive opinions follow past good performance instead of suggesting promising future markets.

2. Literature survey

Forecasting stock market with fuzzy neural networks have been widely used to forecast indices and prices of stock market due to significant properties of treating non-linear data with self-learning capability. However, neural networks suffer from the difficulty to deal with qualitative information and the "black box" syndrome that more or less limited their applications [R.J. Li, Z.B. Xiong, 18-21 August 2005]. To overcome the drawback of neural networks, in this study we proposed a fuzzy

neural network that is a class of adaptive networks and functionally equivalent to a fuzzy inference system. The experimental result based on the comprehensive index of stock market indicate that the suggested fuzzy neural network could be an efficient system to forecast financial time series. Financial prediction is always one of the hottest topics for research studies and commercial applications. With the rapid growth of Internet technology in recent years, e-finance has become an important application of e-commerce. However, in this "sea" of information, available in the Internet, an "intelligent" financial web-mining and stock prediction system can be a key to success. The author proposes the iJADE Stock Advisor—an intelligent agent-based stock prediction system using our proposed hybrid radial basis-function recurrent network (HRBFN). By using 10 years stock pricing information (1990–1999), consisting of 33 major Hong Kong stocks for testing, the iJADE Stock Advisor has achieved promising results in terms of efficiency, accuracy, and mobility as compared with other stock prediction models. Also, various analyzes on this stock advisory system have been performed: including round trip time (RTT) analysis, window-size evaluation test, and stock prediction performance test.

One type of investments in securities is stock as a symbol of ownership for individual or corporation within a company. For making investment strategy, investors should pay more attention to the co-operation of the stock price. Several methods are used to see moving of the stock price, one of them is data mining. This paper proposes Association rule mining algorithm that is used to find the stock movement relationship. Fuzzy approach is used as tools to categorize data based on the sequence of transaction and time. Experimental result shows that the association rule method can solve problems. By using some values, some strong association rules among companies can be obtained.

3. Existing system

In a dynamic environment such as the stock market, the non-linearity of the time series is pronounced, immediately affecting the efficacy of stock price forecasts. If the original load data are directly applied to a train model without eliminating noise, the high-frequency components may disturb the forecasted load patterns. To improve the load forecasting accuracy, a wavelet denoising technique is usually adopted to extract the low frequency component of the load pattern. Therefore, these data are utilized in load forecasting methods. However, many parameters should be determined during the procedure of wavelet analysis, such as the determination of the decomposition layers, wavelet basis function, and threshold function; furthermore, each of the parameters is a large amount and is hard to determine objectively. Because of the subjective selection of the parameters for the wavelet, different researchers obtain various denoising effects; thus, researchers often need to conduct a large amount of data experiments to gain a satisfactory result before continuing the study. However, one of the major advantages of the SSA compared to other approaches is that only two parameters are required to model the time series under analysis. It is relatively new data-driven or non-parametric technique developed to model a nonlinear and/or not stationary as well as noisy short time series. Moreover, it is considered from Claudio that the SSA does not depend on a priori defined functions such as the Fourier approach (based on sine and cosine functions), but it generates a set of components directly from the time series under study. Additionally, the SSA technique can compute periodic or quasi periodic components and a slowly changing trend. To the best of our knowledge, the SSA is not applied to denoise the power load time series. Most of the references predict the components that are decomposed by the SSA and then reconstruct them. In this paper, we denoise the power load by the SSA and then conduct the linear and non-linear methods to validate whether this denoising technique can help the linear and non-linear models to further improve the accuracy of forecasting.

4. Project objectives

The sole purpose of this system is to predict the values of the stock based on the historical data. This help the user in investing his money in the right company. The project is based on an application therefore it can be used on any operating system. The System also helps the user to compare between the stocks and select the stock appropriate to his/her choices. The System includes the following features:

- The system provides a registration system for the user so that the user can keep track of their stocks.
- The system provides an alert system to send alerts related to stocks to the user.
- The system allows user to compare between the stocks.

5. Proposed system

Our system is based on real time data thus it can be used anytime. The system takes historical data from yahoo finance as the input and then uses the back-propagation algorithm to predict the future values of the stock using concept of artificial neural networks. Our system provides features like comparing two or more stocks and helping the user to select best available stock from the selected once. System provides detail analysis regarding the graph of a particular stock. The system also has a feature of sending alert messages to the user regarding the information about the stock.

The hybrid system exhibited outstanding prediction performance and it improves overall profit for investment performance. The proposed model is a promising predictive technique for highly non-linear time series, whose patterns are difficult to capture by traditional models. A recommendation and notification system for people while they getting loss on spending more money on any product. So, in this system we will analyze the current rate of the product and stock. System will recommend the high profitable product when the stock rate

is increase and also it will recommend which product can put on sale to get a profit. Through this we can help the people whenever they face loss on investment, purchase and sale.

6. Architecture diagram

The system takes historical sets from the stock server that is previously updated in the stock products updated list. It is retrieved from the stock server and analyzed using SVM (Support Vector Machine) algorithm. The data sets retrieved using big data and it is processed using map reduce technology.

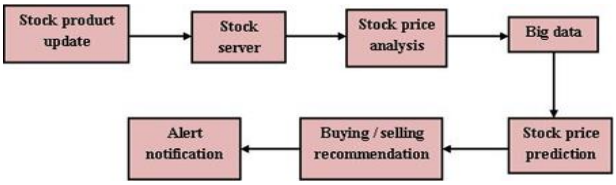

Fig. 1. Architecture diagram

The stock price has been predicted and the buying and selling recommendation has been intimated to the client. The alert notification is provided for the client regarding the fall and rise of future stock values.

7. Module description

- User Interface
- Stock Prediction
- Big Data Analysis & Prediction
- Recommendation
- Alert Notification

A. User interface

An interface is a set of commands or menus through which a user communicates with a program. The user interface is one of the most important parts of any program because it determines how easily you can make the program do what you want. Here we implement a design to get an information from the user likes their personal details and products to sale. User will update what are the products they had on their hand through UI.

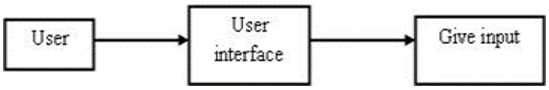

Fig. 2. User interface

B. Stock prediction using SVM algorithm

In this module, we predict the stock market price on the day. Today in economic level, enormous number of stock prediction system is implemented but those information's are not reachable to the common people. So, many business people are getting loss because of improper prediction of stock rate. In our paper we explore the prediction of stock rate and also about product sale.

First, SVM has been rigorously tested in the past for providing higher efficiency than the other classifiers. Second, it is capable of handling the over fitting problem which deals with the appropriate handling of unknown datasets to produce related outputs. In context of theft detection, this unique attribute of SVM plays a significant role in classifying customers as malicious or genuine based on their electric usage profiles. In addition to this, SVM along with different kernels separates the data which are not linearly separable.

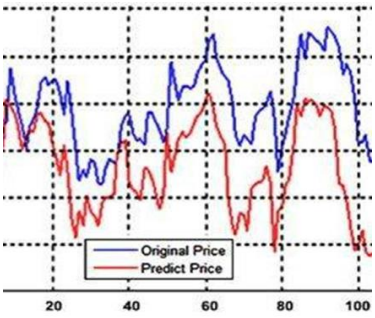

Fig. 3. SVM algorithm

C. Big data analysis and prediction

Big data is the technology to handle huge number of different structures of database. Same way why we are using big data in our system? It is a distributed system to gather the information and execute the information. In our implementation we predict the product rate to the people for user profit.

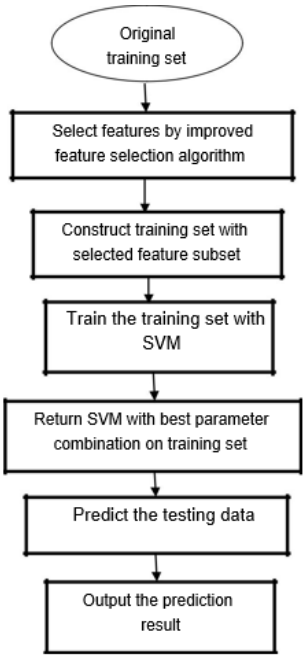

Fig. 4. Data analysis

D. Recommendation and alert notification

In this module, we implement the system to analyze and recommend the product to sale or purchase on profit basis. Because people are buying and selling gold and other properties on everyday basis. But the question is that by doing this whether

they are getting profit or not. So that we recommend the product based on profit. In this module, we notify the people to purchase or sale the profit. Whenever stock rate increases, our system will send notification to the people to sale the product. If decrease we suggest the people to purchase the product. System will recommend the high profitable product when the stock rate is increase and also it will recommend which product can put on sale to get a profit. This project proposes a new prediction algorithm that exploits the temporal correlation among global stock markets and various financial products to predict the next day stock trend with the aid of SVM along with Alert System.

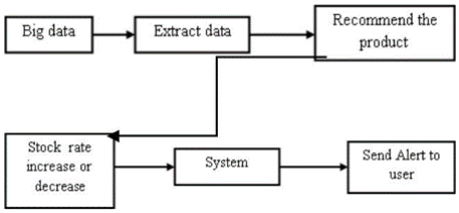

Fig. 5. Alert system

8. Methodology

SVM is the supervised machine learning method which is used for classification of data. The primary goal of using SVM is to classify the unseen data accurately by minimizing the classification error using a decision function. This is achieved by training the SVM on the training data and afterward using it to predict the output class of the unseen data. In this section, SVM-based approach has been put forth for the purpose of theft detection.

A. Data preprocessing

In this step, raw data from various sensors and smart meters deployed in the SG environment is collected and transformed into the format recognized by SVM classifier, categorical attributes in the collected data need to be converted into numeric format.

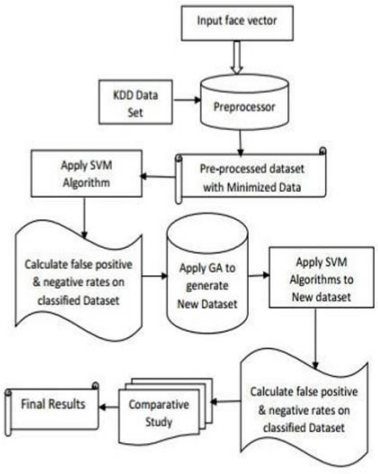

Fig. 6. Data processing

B. Train and test the classifier

After the selection of kernel and modeling parameters, the SVM classifier is trained on the training set. Once the classifier is trained, then it becomes capable of classifying the new test values based on the trained classification model. Apache Hadoop Map Reduce is a framework for processing large data sets in parallel across a Hadoop cluster. Data analysis uses a two-step map and reduce process. The top-level unit of work in Map Reduce is a job. A job usually has a map and reduce phase, though the reduce phase can be omitted.

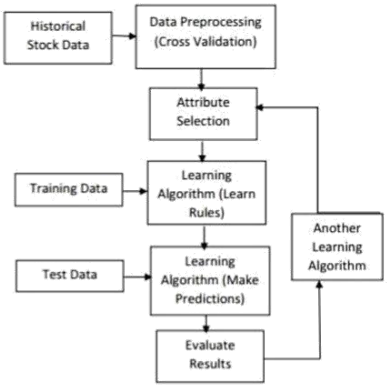

Fig. 7. Classifier

With the exponential growth in the volume of data, the need for its fast retrieval has increasingly become imminent. This requires heavy amount of searching that leads to computational complexities. The scatter technique employs parallelizing the processes that concurrently operate on data in order to provide an efficient search. On the other hand, the gather technique collects data in parallel that leads to fast retrieval of data.

9. Results and discussion

In order to answer our research questions on the effectiveness of discrete stock prediction and the best textual representation; we tested our model against a regression-based predictor using the three dimensions of analysis; measures of Closeness, Directional Accuracy, and a Simulated Trading Engine. Textual data itself can arise from two sources; company generated and independently generated sources. Company generated sources such as quarterly and annual reports can provide a rich linguistic structure that if properly read can indicate how the company will perform in the future (Kloptchenko, Eklund et al. 2004). This textual wealth of information may not be explicitly shown in the financial ratios but encapsulated in forward-looking statements or other textual locations. Independent sources such as analyst recommendations, news outlets, and wire services can provide a more balanced look at the company and have a lesser potential to bias news reports. Discussion boards can also provide independently generated financial news; however, they can b suspect sources.

10. Conclusions and future enhancements

From our research we found that our machine learning model performed much better prediction regarding the stock price than linear regression. These results were consistent throughout the three-evaluation metrics of Closeness, Directional Accuracy, and Simulated Trading across all three textual representations. We further found that Named Entities performed best of the three representations tested. While it did not have the best Directional Accuracy, we feel that there is some room for further optimization. We believe that this representation success arises from its ability to abstract the article in a minimally representative way. Future research includes using other machine learning techniques such as Relevance Vector Regression, which promises to have better accuracy and fewer vectors in classification. It would also be worthwhile to pursue expanding the selection of stocks outside of the S&P500. While the S&P500 is a fairly stable set of companies, perhaps more volatile and less tracked companies may provide interesting results. Lastly, while we trained our system on the entire S&P500, it would be a good idea to try more selective article training such as industry groups or company peer group training and examine those results in terms of prediction accuracy. While the findings presented here are certainly interesting, we acknowledge that they rely on a small dataset. Using a larger dataset would help offset any market biases that are associated with using a compressed period of time, such as the effects of cyclic stocks, earnings reports, and other unexpected surprises. However, datasets of several year duration could be imaginably unwieldy for this type of research.

References

[1] J. A. K. Suykens, "Nonlinear modelling and support vector machines," in *Proc. IEEE Instrum. Meas. Technol. Conf*., 2001, pp. 287–294.
[2] R.J. Li, Z.B.Xiong, "Forecasting Stock Market With Fuzzy Neural Networks," *Fourth International Conference on Machine Learning and Cybernetics*, Guangzhou, 18-21, August 2005.
[3] R. S. T. Lee, "iJADE stock advisor: An intelligent agent based stock prediction system using hybrid RBF recurrent network," *IEEE Trans. Syst., Man, Cybern.. A, Syst. Humans*, vol. 34, no. 3, pp. 421–427, May 2004.
[4] A. K. Rout, R. Bisoi, and P. K.Dash, "A low complexity evolutionary computationally efficient recurrent functional link neural network for time series forecasting," in *Proc. 2015 IEEE Power Commun. Inf. Technol. Conf.*, pp. 576–582.
[5] S. Shen, H. Jiang, and T. Zhang, "Stock market forecasting using machine learning algorithms," Dept. Elect. Eng., Stanford Univ., Stanford, CA, pp. 1–5, 2012.
[6] J.-S. Chou and A.-D. Pham, "Smart artificial firefly colony algorithm based support vector regression for enhanced forecasting in civil engineering," in *Comput. -Aided Civil Infrastructure Eng*., vol. 30, no. 9, pp. 715–732, 2015.
[7] A. A. Arafah and I. Mukhlash, "The application of fuzzy association rule on co-movement analyze of Indonesian stock price," in *Procedia Comput. Sci*., vol. 59, pp. 235–243, 2015.
[8] A. Jindal, A. Dua, K. Kaur, M.Singh, N. Kumar, and S. Mishra, "Decision tree and SVM-based data analytics for theft detection in smart grid," in *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 1005–1016, Jun. 2016.
[9] V. Behbood, J. Lu, and G. Q. Zhang, "Fuzzy refinement domain adaptation for long term prediction in banking ecosystem," in *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1637–1646, May 2014.
[10] J.-S. Chou and N.-T. Ngo, "Engineering strength of fiber-reinforced soil estimated by swarm intelligence optimized regression system," in *Neural Comput. Appl.*, pp. 1–16.
[11] D. Saini, A. Saxena, and R. C.Bansal, "Electricity price forecasting by linear regression and SVM," in *Proc. Int. Conf. Recent Adv. Innov. Eng*., pp. 1–7, 2016.
[12] J. Wang, S. Jin, S. Qin, and H. Jiang, "Swarm intelligence-based hybrid models for short-term power load prediction," in *Math. Probl. Eng*., vol. 2014, 2014.