International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 06 | June 2019 www.irjet.net p-ISSN: 2395-0072

Prediction of Trend in Stock Index using Machine Learning Algorithms with a Special Preference to Trend Deterministic Data

Amitha Nazreen M V[1], Dr. Sabu K[2]

[1]*M. Tech Scholar, College of Engineering, Trivandrum, Kerala, India*
[2]*Associate Professor, Dept. of Mechanical Engineering, College of Engineering, Trivandrum, Kerala, India*

Abstract: *Prediction of stock price index is a highly complicated and very difficult task because there are too many factors such as economic conditions, trader's expectations and other environmental factors that may influence stock prices.*

This paper focus on predicting the direction of movement of stock price index for Indian stock markets. For this, three different models like Support Vector Machine, Random forest and Naïve Bayes Classifier are using inorder to predict the future stock market movement values with accuracy. For that, two approaches for input to these models are also used. The first approach for input data involves computation of ten technical parameters using stock trading data (open, high, low & close prices) while the second approach focuses on representing these technical parameters as trend deterministic data because while using actual values of the stock market movement, accurate prediction values cannot be achieved.

Therefore this paper proposes to provide a special preference to trend deterministic data approach and this methodology will help in achieving better results too. Accuracy of each of the prediction models for each of the two input approaches are also being evaluated in order to validate the performance of these models.

Keywords: Prediction, Stock movement, Stock market, Support Vector Machine, Random forest, Naïve Bayes Classifier.

INTRODUCTION

Stock prediction is termed as the determination of the future market price of a stock or other financial element listed on a stock exchange. For maximizing the investor's gains, one should predict the future value of a stock. Predicting the stock market is not a simple task, mainly as a consequence of the close to random walk behaviour of a stock time series. Stock market plays an important role in the business as well as the economic development of a country. Hence, profitability of investors in the stock market mainly depends in the predictability of stock prices. There are several motivations behind the predictions in stock market prices. The Primary motive is financial gain itself. So if an individual is able to predict it with higher accuracy, he could definitely yield higher returns. Thus many people including researchers, investment professionals and average investors are continuously looking for a superior system which can yield higher returns for them when compared to traditional ones.

Another motivation for analysis during this field is that it possesses several theoretical and experimental challenges. The most important of these is the Efficient Market Hypothesis (EMH), the hypothesis says that in an efficient market, stock prices fully reflect all the available information about its market and its constituents and thus any opportunity of earnings excess profit ceases to exist.

Various technical, fundamental and statistical indicators have been proposed during last decades and many of the investors are still using it to predict the stock market trends. Fundamental and technical analyses were the two methods used to forecast stock prices. First is the fundamental analysis. In this, investors looks at intrinsic value of stocks, performance of the industry and economy, political climate etc. to decide whether to invest or not. While technical analysis is the evaluation of stocks by means of studying statistics generated by market activity, such as past prices and volumes. Technical analysts do not decide to determine a security' s intrinsic price however instead use stock charts to spot patterns and trends which can recommend the future behaviour of the stocks. The most commonly used technique was Artificial Neural Network (ANNs). In most of the cases, over fitting problem arised due to the large number of parameters to fix, and the user uses the knowledge about the important features of the inputs in solving a problem. Also, there is one alternative to reduce these kind of problems, uses Support Vector Machines instead of using the earlier one.

STATEMENT OF PROBLEM

Due to fluctuations in the market, stock market is highly extreme difficult to understand and are unsure in nature. The primary problem in this study is the importance of performance of deep learning and machine learning models in improving the financial stock market analysis and how it helps in predicting Indian stock market index. Hence, if the investor can identify the best opportunities to invest, he could definitely yield higher returns. Good stock investment at bad times can results in disastrous effects than when it is investing in a stock at the correct time can ensure profits.

Today, trading problem was faced by the financial clients or investors as they are not certain on which stocks to buy or which stocks to sell in order to get best optimum output. According to investor's perspective, decision making is very difficult in different stocks in which sector they are belong to. Investor's faced problems in identifying which stocks in which sector perform well as compared to others.

The problem is to be addressed in this study is while using actual values of stock market movement, accurate prediction values cannot be achieved. This work therefore proposes to provide trend deterministic data. This methodology will help in achieving better results. Trend deterministic data can be obtained using technical indicators.

OBJECTIVES

The major objectives of includes:

- To propose 3 different models for predicting stock index movement with the help of technical indicators.
- To evaluate the performance by calculating accuracy and precision of each prediction models.
- To evaluate decision making strategies according to the performance of each prediction models.

LITERATURE REVIEW

It is possible to predict the price of the stock while it is based on trading data and it makes the price of stocks informationally efficient. The stock prices are also affected by factors which are uncertain in nature such as political scenario of the country and public image of the company. If efficient algorithms are used to predict the trend of stock and stock price index then information data obtained from the help of stock prices can be called efficiently pre-processed. Lately, many techniques came into existence which helps in the prediction of stock trends. One of the old techniques are classical regression which used for stock prediction. Stock data can be categorized in the category as time series data of non-stationery nature and also non-linear machine learning techniques are in use. Most popular techniques which are used for predicting stock and stock price index movement are Support vector machines (SVM) and Artificial neural networks (ANN). Both of these algorithms uses their own ideas to learn patterns. ANN uses the technique of functioning as a brain similarly like our brain creates neural networks to function efficiently.

Vapnik (1999) has developed one of the widely used SVM algorithms. It functions by searching for a hyper plane in higher dimensions for different classes. The use of kernel function, scarcity of the solution and the ability to control the decision function makes support vector machine (SVM) one of the special type of learning algorithms.

Abraham et.al (2001) developed a soft computing technique which is hybrid in nature and used for trend analysis and automated stock market forecasting. To predict the stock values it uses neuro-fuzzy system and Nasdaq-100 index with neural network. With the help of these techniques they are able to do stock forecasting one day ahead of the others and it gives promising results for trend prediction and forecasting.

The probabilistic neural network (PNN) was developed by Chen et.al (2003) and it forecast the direction of index with the help of historical information. The study of buy-and-hold strategy in the form of investment strategies gives the direction for examining the empirical results of PNN based investment strategies. It also incorporates the parametric GMM model and random walk model for the functioning of investment strategies.

Most of experiment data shows that SVM performs way better than any other classification methods. To predict the daily price change of stocks in Korea Composites Stock Index (KOSPI), Kim (2003) used SVM and to form the initial attributes they used twelve technical indicators. The comparison of SVM with case-based reasoning and back propagation neural network (BPN) has been done in this study and the results shows better performance of SVM then CBR and BPN. Random forest technique uses samples for creating n classification trees and predicts on the basis of what majority of trees predicts. This trained method of ensembling represents only a single hypothesis and it is not totally necessary that hypothesis space of the models will be able to contain this single hypothesis. It shows us that the ensemble functions can be more flexible than the functions from which they are obtained. This flexibility helps them to totally fit the given trained data which can be useful more than a single model. But when it comes to practise some of the ensemble techniques such as bagging helps to lower the problems which are related to fitting of training data.

Huang et.al (2005) used the method which enables SVM for investigating the predictability of stochastic movement direction and it is done by the weakly movement direction obtained from NIKKI 225 index. Techniques such as Elman Back propagation neural network, Linear Discriminant analysis and Quadratic discriminant analysis are compared with SVM for this study.

The Hang Seng index which is used in the Hong Kong stock market was predicted with the help of total

ten data mining techniques. It enables to predict the price movement of the index. The set of different techniques used for data mining are tree based classification, K-nearest neighbour classification, linear discriminant analysis (LDA), Neural Network, Quadratic discriminant analysis (QDA), Support vector machine (SVM), Naïve-Bayes bsed on kernel estimation, least square support vector machine and Bayesian classification with Gaussian method. Among all these techniques the results shows the better performance of SVM and LS-SVM which are predicting performing better than the other models.

Sun et.al (2012) developed a method of predicting financial distress which is called financial distress prediction (FDP) based on SVM technique. Diversity analysis along with individual performance was considered for designing the algorithm of selected SVM which uses base classifier for individual candidates. The result from the above method shows that SVM is more significant and superior than individual classifier method.

The study by Ahmed (2008) investigated the character of the causative relationships between stock value and the main macro-economic variables representing real real and financial sector of the Indian economy for the period March, 1995-2007 using quarterly data. The study discovered that the movement of stock value was not solely the end result of behaviour of key macro-economic variables however it had been additionally one of the reason of direction in other macro dimension in the Indian economy.

The above discussion about different algorithms and techniques shows different ways of tackling the problem. It also shows that each algorithm and techniques comes with its own advantages and disadvantages. So the final prediction is highly influenced by the algorithm used for the prediction and the technique adopted. It ids also affected by the way inputs are represented in the problem. The method of segregating important features of algorithms and making use of them as input instead of using all the available features can help in the improvement of the prediction accuracy of the model.

METHODOLOGY

The data employed in this study consists of daily prices of BSE S&P Index. The dataset contains the trading days ranging from February 2011 to 2018. The data collection is done from www.investing.com website's historical data. In this study, 3 different prediction models are planning to design for predicting stock index for a certain period.

This paper aims to focus on comparing the performance of the prediction models such as SVM, random forest and Naïve Bayes classifier for the task of predicting stock index movement. Before prediction, technical analysis is done accordingly by generating inputs to the models. These inputs are generated by the calculation of 10 technical indicators by the addition of several variables from the data given. These calculations are done in order to categorize the value as trend deterministic data which are used for model processing for attaining output. Several technical indicators such as moving averages, momentum, stochastic, relative strength index, moving average convergence divergence, William % R, commodity channel index, and aroon function are used for the preparation of trend deterministic data. The values obtained from the calculations are represented as +1 or -1 (down or up trends) in order to create the input for modelling. These inputs are given to the models and corresponding outputs are produced by future prediction by taking the past movement into consideration. Validations of models are evaluated by performance calculation that is being done by using appropriate tools.

TECHNICAL ANALYSIS: INPUT CONTRIBUTION

There are two types of analysis which investor seeks before investing in a stock, i.e fundamental analysis and technical analysis. Here, technical analysis is the evaluation of stocks by means of studying statistics generated by market activity, such as past prices and volumes. Technical indicators are the key elements in a technical analysis. This is done by a technical analysts, one who do not attempt to measure a security's intrinsic value but instead of that he/she use stock charts to identify patterns and trends that may suggest how a stock will behave in the coming years. In this, 10 technical indicators are selected to perform the analysis under appropriate selection criteria according to the data required. These technical indicators are used to generate the input form modelling. For input generation, two methods of representation of data are applied. The first approach uses continuous values representation. The second one uses Trend Deterministic representation for the inputs.

Using these indicator values, the trend deterministic input set is prepared and given to the predictor models. Performance of all the models under study is evaluated also for this representation of inputs.

MODEL IMPLEMENTATION

Three prediction models are used in this study. They are SVM, random forest and Naïve Bayes classifier with their respective algorithms. SVM is a deep learning model which is generally used for small dataset. So, it can be used in this work too. Several researchers recommended to use random forest, which gives higher prediction values by modelling. While comparing the three, random forest outperforms the other two in

related works. At last, performance validation is also done to determine which model is better accurate than others.

PERFORMANCE VALIDATION

Accuracy and f-measure are used to evaluate the performance of predicted models. By this, confusion matrix is developed in order to validate the accuracy of modelling.

Confusion matrix is table that is often used to describe the performance of a classification model (or classifier) on a set of test data for which the true values are known. It tells about the summary of prediction results on a classification problem. In the field of machine learning, it is also known as "error matrix" because it checks the quality of the output obtained of a problem. That's why it is used as a performance measure in machine learning algorithms.

The theory of validation confess about that the purpose of experiments on comparison data set is to compare the prediction performance of these models for best parameter combinations reported during parameter setting experiments. During this comparison experiment, each of the prediction models is learnt based on best parameters reported by parameter setting experiments.

RESULTS

Summary statistics is prepared based on the result obtained from the calculation of technical indicators. It consists of maximum value, minimum value, mean, and standard deviation of the calculated technical indicator values.

Three model implementations are done according to the data given and suitable results are obtained. Out of these, 70% of the data is given for training the data set and remaining for testing. Results showed that the test accuracy of Naïve-Bayes classifier algorithms is more accurate when compared with random forest and SVM when applied to BSE S&P sectoral indices of Indian stock market and is expressed in percentage accuracy.

Out of these three algorithms formulated we can see that naïve-bayes classifier outperforms the other two in many of the sectors. This is clearly explained with the comparison between training and testing of data set. So, naïve-bayes classifier gives approximate accuracy when compared with the training accuracy. This is because by training the data set we have to achieve the test accuracy approximate to the training accuracy. That's why naïve-bayes classifier predicts in higher accuracy in most of the sectoral indices.

CONCLUSIONS

The task focused in this paper is to predict direction of movement for stock price indices. Prediction performance of three models namely SVM, random forest and Naive-Bayes is compared based on seven years (2011–2018) of historical data of BSE S&P Sensex from Indian stock markets. Ten technical parameters reflecting the condition of stock price index are used to learn each of these models. Experiments with continuous-valued data show that Naive- Bayes (Gaussian process) model exhibits higher performance with 86% accuracy, SVM exhibits a performance with 36% accuracy and random forest with least performance of 40% accuracy.

In earlier researches, the technical indicators were used directly for prediction while this study first extracts trend related information from each of the technical indicators and then utilizes the same for prediction, resulting in significant improvement in accuracy. The proposal of this Trend Deterministic Data Preparation Layer is a distinct contribution to the research.

REFERENCES

[1] Abraham, A., Nath, B., & Mahanti, P. K. (2001). Hybrid intelligent systems for stock 726 market analysis. *In Computational science-ICCS 2001* (pp. 337–345). Springer.

[2] Garg, A., Sriram, S., & Tai, K. (2013). Empirical analysis of model selection criteria for 737 genetic programming in modeling of time series system. In 2013 *IEEE Conference 738 on Computational Intelligence for Financial Engineering & Economics (CIFEr)* 739 (pp. 90–94). IEEE.

[3] J. Sharmila Vaiz, Dr M. Ramaswami (2016), Forecasting Stock Trend Using Technical Indicators with R, International Journal of Computational Intelligence and Informatics, Vol 6: No. 3.

[4] Kara, Y., Acar Boyacioglu, M., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul stock exchange. *Expert systems with Applications,* 38, 5311–5319.

[5] Kim, K., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. Expert Systems with Applications.

[6] Kim, S. H., & Chun, S. H. (1998). Graded forecasting using an array of bipolar predictions: Application of probabilistic neural networks to a stock market index. International Journal of Forecasting.

[7] Khemchandani, R., & Jayadeva Chandra, S. (2009). Knowledge based proximal support vector machines. European Journal of Operational Research.

[8] Karaatli, M., Gungor, I., Demir, Y., & Kalayci, S. (2005). Estimating stock marketmovements with neural network approach. Journal of Balikesir University.

[9] Md. Rafiul Hassan, Baikunth Nath, Michael Kirley (2007), A Fusion Model of HMM, ANN and GA for Stock market Forecasting, *Expert System with Applications*, 33, 171-180.

[10] Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and 762 empirical work/. *The Journal of Finance*, 25, 383–417.

[11] Manish, K., & Thenmozhi, M. (2005). Forecasting stock index movement: A comparison of support vector machines and random forest. In Proceedings of ninth Indian institute of capital markets conference, Mumbai, India.

[12] Rajat Singla (2015), Prediction from Technical Analysis-A Study of Indian Stock Market, International Journal of Engineering Technology, Management and Applied Sciences, Volume 3, Issue 4, ISSN 2349-4476.

[13] Saad, E. W., Prokhorov, D. C., & Wunsch, D. C. (1998). Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. IEEE Transactions on Neural Networks.

[14] Wei Huang, Yoshiteru Nakamori, Shou-Yang Wang (2005), Forecasting Stock Market Movement Direction with Support Vector Machine, *Computers and Operations Research*, 32, 2513-2522.