

MACHINE LEARNING APPROACH IN STOCK MARKET PREDICTION

Raut Sushrut Deepak ¹, Shinde Isha Uday ², Dr. D. Malathi ³

^{1,2}B.Tech Student, ³Professor

Department of Computer Science and Engineering SRM University, Kattankulathur, Chennai.

sushrutraut94@gmail.com¹, eshashinde31@gmail.com², malathi.d@ktr.srmuniv.ac.in³

Abstract: High level of accuracy and precision is the key factor in predicting a stock market. The technical, fundamental or the time series analysis is used by most of the stockbrokers while making the predictions. Nevertheless, these methods cannot be trusted fully, so there is a necessity to provide the supportive method for stock market prediction. In this paper, we propose a Machine Learning (ML) approach that will be trained from available stocks data, gain intelligence and then uses the acquired knowledge for accurate prediction. After the through research of various algorithms and their fitness for different problem domains, Artificial Neural Network (ANN) was found to be the most practical consideration. Neural network models having the features and customisable parameters makes it possible to implement wide number of features along with the crossvalidation sets. The main significant approach, used in this paper for the predicting result is a concept of machine learning and result tested on the Bombay Stock Exchange (BSE) index data set.

To seize the best accurate output, the approach decided to be implemented is machine learning along with supervised classifier. Results are tested on the binary classification done using SVM classifier with a different set of a feature list. Most of the Machine Learning approach for solving business problems have their privilege over statistical methods that do not include AI, although there is an optimal methodology for certain problems. A comparative analysis of selective application is conducted which concluded that ANNs are extensively implemented in forecasting stock prices, returns more control degree of a relative similarity in predicted result.

Keywords: Machine Learning, Stock Market, Artificial neural networks, Bombay Stock Exchange, Support vector machine.

1. Introduction

The trend in stock market prediction is not a new thing yet this issue is kept being discussed by the various organisation. Being able to predict accurately the future financial outcome is equivalent to earning big money. This paper aims at analysing this problem in an academic way which provides a different way of prediction on the market trend. This paper will develop a financial data predictor program there will be dataset storing all historical stock prices and data will be treated as training sets for the program. To predict the stock prices in BSE, this project aims at developing a program

which serve best solution for accurate predicted stock result.

The parameters like Moving Average Convergence-Divergence (MACD), Stochastic were included in trading rule. This paper is based on clustering approach for designing a trading system on the basis of the historical data instead of identifying redundant behaviour which represents a probable risk for traders using conventional algorithm. Deciding right features is an integral element for the processing and traders should determine features that co-relates to changing the price of given security for example, experiment with different indicators and see if any seem to co-relate with major market term.

There are various functions that are used to predict the parameters. Mainly include, binary threshold, linear threshold, sigmoid and tan hyperbolic. In this approach, with the use of supervised learning classifier to predict stock price movement based on financial index report, and evaluate their potency is proposed. Statistical analytic methods in financial market have become stock modelling. In spite of many benefits, there are limitations that should be investigated such as relevance of the result, best topology for the certain problems and its validity along with accuracy in near future. Proposed model in this paper can be a acceptable solution for stock prediction.

A neural networks based model has been used in predicting of the stock market for an NASDAQ's stock value with a given input parameters of share market[1]. The value of NASDAQ Stock Market index has been used in Real exchange rate. Mizuno and friends had implemented the Tokyo stock exchange to forecast buying and selling signals with an overall accuracy rate of 63% by using ANN. Sexton and friends initiated with learning at random points that notify in the training process. Phua and friends had implemented ANNs with the genetic algorithm to the stock market value of Singapore and forecast the market value with an forecasting rate of 81%.

Performance of Hierarchical clustering algorithm and reverse K means is compared on the basis of class wise cluster building ability of algorithm[2]. The prediction on the short-term stock prices movement is done by an effective clustering method, HRK (Hierarchical agglomerative and Recursive K-means clustering) after the release of financial reports.

A hybridised approach [3] is used to improve existing approaches by combining the variables of fundamental and technical analysis of stock market for prediction of future price of stock. The research

implement three-layer (one hidden layer) multilayer perceptron model (a feed-forward neural network model) trained with back propagation algorithm. The actual results obtained showed high level of accuracy for daily stock price prediction with hybridised approach executes better than technical analysis approach.

As per the research evidence, the random walk (RW)[4] is also an effective model for forecasting data related to finance. Another promising alternative is Artificial neural network having unique capability of self-adaptive modelling. A new hybrid technique which uniquely combines the RW, FANN is proposed to predict the financial data. Actual results from real world financial time series clearly suggest that hybrid method substantially improved the overall predicting accuracies and also dominates each of the individual component models.

Market performance of Karachi Stock Exchange (KSE)[5] on day closing using different machine learning techniques. The forecasting model apply distinct attributes for the input and predicts market as Positive & Negative. The attributes applied in the model involve Oil rates, Gold & Silver rates, Interest rate, Foreign Exchange (FEX) rate, NEWS and social media feed. Following are different factors which have effect on market performance in different studies- Market history, the news, general public mood, commodity price, interest rate, Foreign Exchange. The outcome of the research confirms that machine learning techniques are efficient in predicting the stock market performance. Karachi Stock Market with KSE-100 index follows a pattern that can be predicted using machine learning techniques. The Multi-Layer Perceptron algorithm of machine learning predicted 77% correct market performance.

Swarm Intelligence optimisation technique named Cuckoo search[6] is most simple to accommodate the parameters of SVM. The proposed hybrid CS-SVM technique has been demonstrated capability to produce more accurate results in comparison to ANN and in the forecasting of the stock price movement. Forecasting activity is segregate in three groups of short, medium and long forecasting. Short term forecasting interval comprises few minutes, hour or days between a week. The forecasting interval beyond one week and within a month symbolise the medium term forecasting while long term may cover the period of one to few years. The forecasting accuracy of ANN SVM and CS-SVM for Indian stock market was analysed. The gained results

specify that the CS-SVM method is able to obtain higher accuracy rate in comparison to regular ANN method. The output depicts the accuracy improvement of CS-SVM on 80.23% - 82.2% for BSE-Sensex.

Prediction future stock price uses parse records then calculate predicted value and send to user and automatically perform operations like purchase and sale shares using Automation concept. Naïve Bayes Algorithm is used[7]. Real time Access for downloading log forms yahoo finance website is taken and store in dataset. The probable market prediction target can be the future stock price or the volatility of the prices or market trend. In the prediction there are two types like Dummy and Real time prediction used in stock market System. In Dummy prediction they have define some rules and predict the future price of shares by calculating average price. In the Real time prediction compulsory used internet and saw current price of shares of companies.

2. Problem Statement

This paper consists of three modules to check the accuracy of the result; the result are predicted on the next week, next day and next minute stock values.

Trading in shares is big business in many economies including BSE. Currently, Stockbrokers who execute trades and advise clients, rely on their experience, technical analysis (price trends) or fundamental analysis (buy and hold) in picking their stocks. These current methods are subjective and are usually short sighted due to their limited capacity. With the value of trade money involved, improper investment could easily mean great losses to investors, especially if they keep making wrong decisions. Lack of guaranteed returns has also led to the reluctance by potential investors to participate in the market. It is therefore desirable to have a tool that can guide on the most likely next day prices (prediction) as a basis of making any investment decision.

The use of fundamental and technical analysis methods are basis of the predictions of future stock price movement. These tools show a trend on future movement and not the figure of the most likely trade price for any stock in future. It is therefore desirable to have a tool that does not just point at a direction of price movement. Machine Learning methods that can actually analyse the stock prices over time and gain intelligence, then use this intelligence in prediction, can be used to model such a tool.

<i>Authors</i>	<i>Methodology/ Algorithm</i>	<i>Dataset used</i>	<i>%of accuracy</i>	<i>Advantages/ Results</i>	<i>Limitations</i>	<i>Future Enhancement</i>
[1]	ANN, MLP, Feed Forward, Back Propagation	NAS DAQ stock market index	81	It eases an approximation of any input or output map	Data train very slowly and requires lots of training data	Determine the critical impact of specific fundamental analysis

[2]	K-menas, Hierarchical clustering algorithm and reverse K means	Yahoo, BSE,NSE financial databases	-	Consider qualitative and quantitative features	-	-
[3]	ANN	Daily Stock Prices	-	The hybridised approach to enhance the quality stock prediction	-	Prediction for Investor
[4]	A N N,Random Walk() RW, feed forward A N N and Elman ANN	USD-INR exchange rate GBP-USD exchange rate,S&P 500, IBM stock price	Error mean value 0.571	Hybrid method substantially improve d the overall forecasti ng accuracies	-	-
[5]	Single Layer Perceptron,LP, RBF and SVM	KSE -100 Index	77	Performance of the model is good even with lack of resources	The most related attribute to the KSE performance was found to be the petrol prices	-
[6]	Improve traditional B P algorithm	Chinese stock market	25% improvement	Exceeded convergence speed,prediction, precision and processing	-	-
[7]	Naive Bayes Algorithm	Stock data from yahoo finance	-	Automatic stock purchased model can be done using this algorithm	Limited to favourable outcome of buy and sell does not show stock analysis	Useful to develop web portal for automatic trade

[8]	Linear Regression and Neural Network Standard FFB	Four years of data from indexes in the Iran Tehran Stock Exchange	-	Proved more accurate result by ANN techniques over the regression mode	Applicable to the Single trading company	Feed forward model can be enhanced with more training pattern
[9]	Feed forward multilayer perceptron with error back propagation	NSE compile in five year period 2008-2012	MAPE of between 0.77% and 1.91%	ANN based model can be used to develop low RMSE system	-	ANN Techniques can be used in stock market software

3. Methodology

In this paper the stock market prediction is done by the Support Vector Machine along with RBF kernel algorithm.

A. Support Vector Machine

SVM is considered to be one of the most suitable algorithms available for time series prediction. This supervised algorithm can be used in both, regression and classification. The SVM involves plotting of data as point in the space of n dimensions.

These dimensions are attributes that are plotted on particular co-ordinates. SVM algorithm draws a boundary over the data set called the hyper-plane, which separates data into two classes as shown in the Fig 1.

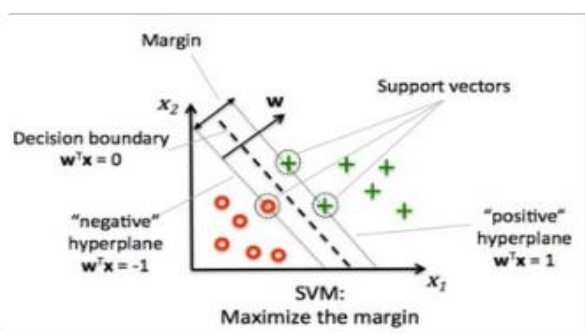


Figure 1. The Support Vector Machine Decision Boundary

The hyper-plane is the decision boundary which is later extended or maximised on either side between the data points.

Considering the same figure, if μ is some unknown data point and w is a vector which is perpendicular to the hyper-plane, then the SVM decision rule will be

$$w^T \mu + b \geq 0 \quad (1)$$

The width w of the hyper-plane must be maximised to increase the spread

$$W = [2 / \|w\|] \quad (2)$$

$$W = \max[2 / \|w\|] \quad (3)$$

Applying Lagrange's multiplier as

$$L = 0.5 \|w\|^2 \rightarrow -\sum \alpha_i [y_i (w^T x_i + b) - 1] \quad (4)$$

$$L = \sum \alpha_i - 0.5 \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i x_j \quad (5)$$

The updated decision rule will be

$$(\sum \alpha_i y_i x_i) \mu + b \geq 0 \quad (6)$$

B. RBF

This type of neural network is supervised algorithm and is a feed-forward network. It depends only on radial distance from a point. Like MLP, the RBF networks also have three different layers including input, output and hidden layer. Each hidden neuron represents a single radial basis function which has its centre and width (spread). RBF transforms the Nonlinearly separable classes into the linearly separable classes. The non-linear transfer function ϕ of RBF is given below. Here, the σ is a parameter called the spread, which indicates the selectivity of the neuron.

$$\phi_{ij} = e^{-\frac{\|x - x_j\|^2}{2\sigma^2}} \quad (7)$$

Weights are compared by means of the pseudo-inverse method.

$$w_i \varphi_i(\|x_i - t_i\|) + \dots + w_{m1} \varphi_{m1}(\|x_i - t_{m1}\|) = d_i \quad (8)$$

This can be written in matrix form as;

$$\begin{pmatrix} \varphi_1(\|x_1 - t_1\|) & \dots & \varphi_{m1}(\|x_1 - t_{m1}\|) \\ \vdots & & \vdots \\ \varphi_1(\|x_N - t_1\|) & \dots & \varphi_{m1}(\|x_N - t_{m1}\|) \end{pmatrix} [w_1 \dots w_{m1}]^T = [d_1 \dots d_N]^T \quad (9)$$

$$[w_1 \dots w_{m1}]^T = \Phi^+ [d_1 \dots d_N]^T \quad (10)$$

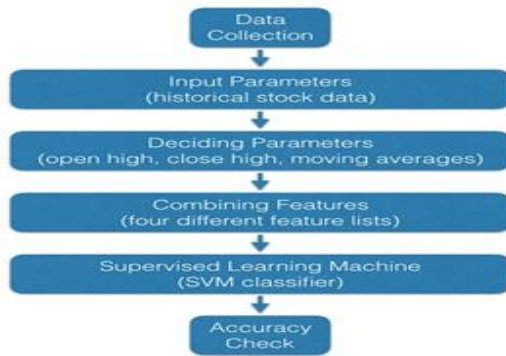


Figure 2. Architecture of the model.

4. Results

Result of the model is calculated on the stock data of the BSE listed companies. Dataset used in prediction is obtained live data from yahoo finance Url and stored data of 2014-2016. Tested on the four companies dataset. Graph is plotted showing prediction accuracy for four different feature list. Results showed that the accuracy upto 89% is achieved.

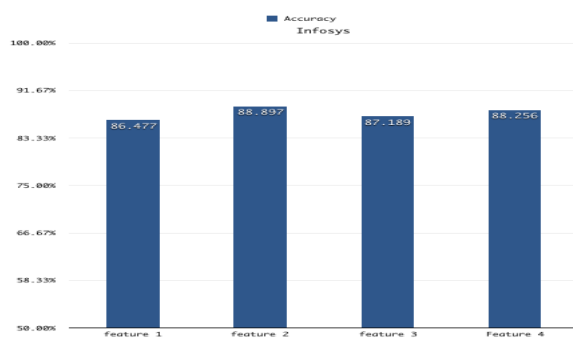


Figure 3. Daily Prediction Accuracy on Infosys

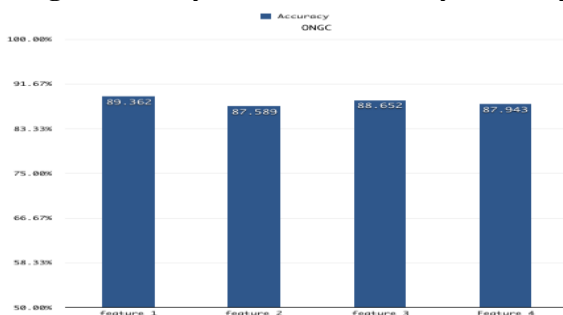


Figure 4. Daily Prediction Accuracy on ONGC

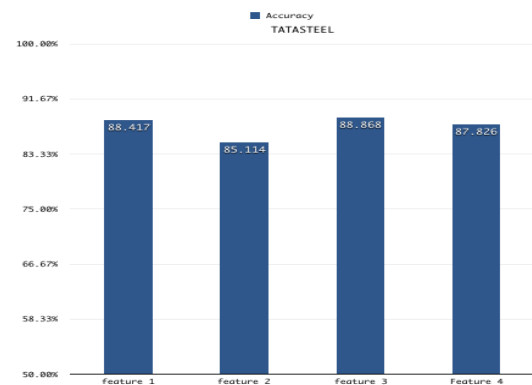


Figure 5. Daily Prediction Accuracy on TATASTEEL

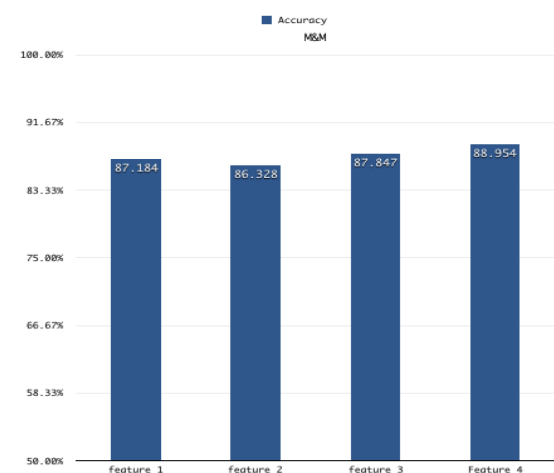


Figure 6. Daily Prediction Accuracy on M&M

SVM - Features 1:
Prediction Accuracy: 82.7956989247

SVM - Features 2:
Prediction Accuracy: 84.9462365591

SVM - Features 3:
Prediction Accuracy: 77.4193548387

SVM - Features 4:
Prediction Accuracy: 80.6451612903

Output Screenshot

5. Conclusion

Based on the results shown in fig(3-6) and experiments performed, it is evident that input data plays an important role in prediction along with machine learning techniques. As is seen in result we have combined the dataset and formed four set of feature list, provided, when we split the data into train and test, the number become very less which is nothing but noise and can be totally removed from the dataset by using filtering techniques and hence the model will be available to predict the outcome much better with absence of noise. Moreover, SVM has proven that we can generator more custom feature set and obtain predictions with great efficiency. We have conducted tests using non linear RBF kernel which showed considerable accuracy in outcome. Most importantly, the above experiment not only helped us in predicting the outcome but also gave us valuable insights about the nature of data, which can be used in future to train our classifiers in a much better way.

The project can be expanded further by improvising feature list and with different classifier. Future work include the use of unsupervised preprocessor along with the supervise classifier.

Reference

- [1]Yunus Yetis; Halid Kaplan; Mo Jamshidi (2014), "Stock Market Prediction by Using Artificial Neural Network", 2014 World Automation Congress (WAC), pages: 718 - 722, DOI: 10.1109/ WAC. 2014.6936118.
- [2]M.Suresh Babu, Dr. N.Geethanjali, Prof B.Satyanarayana (2012), "Clustering Approach to Stock Market Prediction", Int. J. Advanced Networking and Applications , Volume: 03, Issue: 04, Pages: 1281-1291.
- [3]Adebisi Ayodele A., Ayo Charles K., Adebisi Marion O., and Otokiti Sunday O. (2012), "Stock Price Prediction using Neural Network with Hybridized Market Indicators", Journal of Emerging Trends in Computing and Information Sciences, Volume: 03, Issue: 01, Pages: 1-9.
- [4]Ratnadip Adhikari, R. K. Agrawal (2013), "A combination of artificial neural network and random walk models for financial time series forecasting", Springer-Verlag London, DOI 10.1007/ s00521-013-1386-y.
- [5]Mehak Usmani; Syed Hasan Adil; Kamran Raza; Syed Saad Azhar Ali "Stock Market Prediction Using Machine Learning Techniques" 2016 3rd International Conference on Computer and Information Sciences (ICCOINS) Pages: 322 - 327, DOI: 10.1109/ICCOINS.2016.7783235
- [6]Sihua Chen; Changqi Tao; Wei He "New Algorithm of Neural Network and Prediction in China Stock Market" 2009 PacificAsia Conference on Circuits, Communications and Systems 2009 Pages: 686 - 689, DOI: 10.1109/PACCS.2009.140.
- [7]Mahajan Shubhrata D., Deshmukh Kaveri V., Thite Pranit R., Samel Bhavana Y., Prof. Chate P.J (2016), "Stock Market Prediction and Analysis Using Naïve Bayes", International Journal on Recent and Innovation Trends in Computing and Communication, Volume – 4 Issue – 11, Issue Number - Issue: 11121-124.
- [8]Reza Aghababaeian, Tamanna Siddiqui, NajeebAhmadKhan (2011), "Forecasting the Tehran Stock Market by Artificial Neural Network", (IJACSA) International Journal of Advanced Computer Science and Applications, Special Issue on Artificial Intelligence, Issue No: 10.14569, Volume: 010, Page: 303.
- [9]Wanjawa, Barack Wamkaya(2012), "A Neural Network Model for Predicting Stock Market Prices at the Nairobi Securities Exchange", University of Nairobi research archive, Volume – 73404, Page- 56.
- [10] Radu Iacomin (2015), "Stock Market Prediction", 2015 19th International Conference on System Theory, Control and Computing (ICSTCC), October 14-16, Cheile Gradistei, Romania, Pages: 200 - 205, DOI: 10.1109/ICSTCC.2015.7321293.
- [11] R.Kandhan, R.Ananthan, P.Tamilselvan & Mr.D.Sathyamurthy, "An Efficient Privacy-Preserving Search System Over Encrypted Cloud Data", International Innovative Research Journal of Engineering and Technology, pp.27-29, 2017.

