Stock Market Trend Prediction using Machine Learning

A.Akash[1], R.Shanthi[2], R.Aravinth[3], V.Kurunji vendhan[4], D.Veerapandi[5]

UG Student, Dept. of Information Technology, SRM Valliammai Engineering College, Chennai, Tamilnadu, India[1,3,4,5]

Asst. Professor, Dept. of Information Technology, SRM Valliammai Engineering College, Chennai,

Tamilnadu, India[2]

ABSTARCT: Stock market prediction is the 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 stock broker while making the predictions. Nevertheless, these methods cannot be trusted fully, so there is the necessity to provide supportive method. In the system, SVM is implemented for the stock market prediction, which is inefficient. So, In our project we proposed Least Square Support Vector Machine (LS-SVM) and Particle Swarm Optimization (PSO) algorithm is used. The PSO algorithm is employed to optimize LS-SVM to credit the daily stock prices. Proposed is based on the study of stock historical data and technical indicators. PSO algorithm selects best free parameters combination for LS-SVM to avoid over-fitting and local minimization problem and improve prediction accuracy. The proposed model was applied and evaluated using thirteen benchmark financial datasets and comparted with artificial neural network with Levenberg -Marquardt (LM) algorithm. The obtained results showed that the proposed model has efficient prediction accuracy and the potential of PSO algorithm optimizing LS - SVM.

KEYWORDS: Least Square Support Vector Machine, Particle Swarm Optimization Technical Indicator, level of accuracy.

I. INTRODUCTION

Stock price prediction has been at focus for years since it can yield significant profits. the stock market prediction is not a simple process, mainly as a consequence of the close to random- walk behaviour of a stock time series. Two methods are used for forecast stock prices such as Fundamental and technical analyses . most commonly used technique is Artificial Neural Network[1]. ANN's affect by over-fitting problem due to the large number of parameters to fix, and the little prior user knowledge about the inputs are the analysed problem . Also, Support vector machines (SVMs) can be used as an alternative for avoids such a over fitting problem . SVM can be used for compute globally optimal solutions, which tend to fall into local minima. Least squares –support vector machines (LS-SVM) method was presented in , which was reformulated the traditional SVM algorithm[2]. LS-SVM have equality constraints with regularized least square function. Although LS-SVM reduces the SVM procedure, the regularization parameter and the kernel parameters play an important role . Therefore, To selecting a LS-SVM free parameters there is need to establish the methodology in a proper manner. By this way the regression obtained by LS-SVM must be strong against noisy conditions, and it does not need priori user knowledge to learn about the free parameters values in the problem studied. The LS-SVM have a advantages as optimization methods motivated made the researchers to include such stochastic methods in the context of optimizing SVM, Particle swarm optimization (PSO) is one of the most used Evolutionary Algorithms (EAs)[2]. The difference between particle swarm optimization concept and the evolutionary computing is that potential solutions through hyperspace are moving toward "better" solutions, The potential solutions are get by the evolutionary computation schemes which are represented as locations in hyperspace . The optimizer in the particle swarm optimization algorithm, the adjustment of "local" and "global" particles, is similar tothe crossover operation performed by genetic algorithms .The results showed that model has better prediction accuracy and the need of PSO algorithm in LS-SVM optimization.

II. RELATED WORK

Inter-comparison of Artificial Neural Network Algorithms for Time Series Forecasting: Predicting Indian Financial Markets

Shilpa Amit Verma Thadomal ,G. T. ThampiThadomal et al [1] proposed a neural networks based model have been used in predicting of the stock market. One of the methods, as an intelligent data mining, is artificial neural network (ANN). In this paper represents how to predict a NASDAQ's stock value using ANNs with a given input parameters of share market. Real exchange rate value of NASDAQ Stock Market index is used. This paper makes use generalized feed forward networks. The network was trained using input data of stock market price in between 2012 and 2013. It shows a good performance for NASDAQ stock market prediction.

Clustering Approach to Stock Market Prediction

M.SureshBabu, Dr.N.Geethanjali et al [2] proposed clustering is an adaptive procedure in which objects are clustered or grouped together, based on the principle of maximizing the intra-class similarity and minimizing the inter-class similarity. Various clustering algorithms have been developed which results to a good performance on datasets for cluster formation. This paper analyze the major clustering algorithms: K-Means, Hierarchical clustering algorithm and reverse K means and compare the performance of these three major clustering algorithms on the aspect of correctly class wise cluster building ability of algorithm. An effective clustering method, HRK (Hierarchical agglomerative and Recursive K-means clustering) is proposed, to predict the short-term stock price movements after the release of financial reports.

Stock Market Prediction Using Machine Learning Techniques

Mehak Usmani, Syed Hasan Adil et al [3] proposed the main objective of this research is to predict the market performance of Karachi Stock Exchange (KSE) on day closing using different machine learningtechniques. The prediction model uses different attributes as an input and predicts market as Positive & Negative. The attributes used in the model includes Oil rates, Gold & Silver rates, Interest rate, Foreign Exchange (FEX) rate, NEWS and social media feed. The old statistical techniques including Simple Moving Average (SMA) and Autoregressive Integrated Moving Average (ARIMA) are also used as input. The machine learning techniques including Single Layer Perceptron (SLP), Multi-Layer Perceptron (MLP), Radial Basis Function (RBF) and Support Vector Machine (SVM) are compared. All these attributes are studied separately also. The algorithm MLP performed best as compared to other techniques.

A Neural Network Model for Predicting Stock Market Prices at the Nairobi Securities Exchange

Wanjawa, Barack Wamkaya et al [4] proposed Artificial intelligence (AI) model that trains from available stocks data, gains intelligence and then uses the acquired knowledge for predicting future prices. The research singled out Artificial Neural Network (ANN) as the basis for the model, after considering various algorithms and their suitability for different problem domains. Through experimentation, the project developed an ANN model, based on feed forward multi-layer perceptron (MLP) with error back propagation. The final model was of configuration 5:21:21:1 i.e. 5 inputs, with two hidden layers each having 21 neurons and 1 output. To test the model, the research developed a prototype, based on C# programming language and tested it on data of daily trades at the NSE compiled in the five-year period 2008-2012. The holdout method was used for training and testing, with 80% data for training and the balance 20% for testing. The results showed that the model was able to predict the future trends of three chosen stocks correctly, with a Mean Absolute Percentage Error (MAPE) of between 0.77% and 1.91%

Stock Market Prediction and Analysis Using Naïve Bayes

Mahajan Shubhrata D, Deshmukh Kaveri V., Thite Pranit R., Samel Bhavana Y [5] proposed The stock market is the most popular investing places for users. Because of its expected high profit. Recently forecasting stock market returns gaining more attention. The prediction of stock markets is regarded as a challenging task. Data analysis is the way of predicting future value. if future stocks prices will increase or decrease. The main objective of this paper is to predict future stock price using prediction concept. In that Parse Records then calculate predicted value and send

to user. And automatically perform operations like purchase and sale shares using Automation concept. For that use Naïve Bayes Algorithm. There is Real time Access by Download log forms yahoo finance website and Store in dataset.

III. EXISITING SYSTEM

Machine Learning (ML) approach is used to train available stock data, gain intelligence and then uses the acquired knowledge for accurate prediction. After the thorough 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 customizable parameters makes it possible to implement wide number of features along with the cross validation sets. The main significant approach,usedinthispaperforthepredictingresultisa concept of machine learning and result tested on the Bombay Stock Exchange (BSE) index dataset. Different set of feature list are used on the results which is tested on the binary classification. Most of the Machine Learning approach used for business problems have their own statistical methods that do not include AI, although there is an methodology for certain problems which is optimal. Prediction of the future financial outcome is equivalent to making big money. This paper aims at analyzing this problem in an academic way which provides a different way of prediction on the market trend. This financial data predictor program there will be dataset storingallhistoricalstockpricesanddata develop from the paperwillbesuitable to act as training sets for the program. There are many functions that are used to detect the parameters. It include, binary threshold, linear threshold, and tan hyperbolic. In this method , we use supervised learning classifier which predict stock price movement based on financial index report, and estimates their potency Statistical analyticmethodsinfinancialmarkethavebecomestock modelling. Hierarchical clustering algorithm and reverse K means performance is compared on the class wise cluster building algorithm. A hybridized approach 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 short-term stock prices movement prediction is done. To predict the stock pricesinBSE,thisproject aims at developing a program which predict the result.

IV. PROPOSED SYSTEM

The Proposed mode is based on the study of stocks historical data and technical indicators. PSO algorithm selects best free parameters combination for LS-SVM to avoid over-fitting and local minima problems and improve prediction accuracy. The model was evaluated in thirteen benchmark financials datasets that are compared with artificial neural network. The obtained results process that the proposed model has better prediction accuracy and the potential of PSO algorithm in optimizing LS-SVM. The two methods used to forecast stock prices that are Fundamental and technical analyses. . ANN's affect by over-fitting problem due to the large number of parameters to fix, and the little prior user knowledge about the relevance of the inputs in the analysed problem, Least squares –support vector machines (LS-SVM) method was presented in which was reformulated the traditional SVM algorithm. LS-SVM uses equality constraints for a regularized least squares function .LS-SVM are a set of related supervised learning methods hat analysis and optimize data which are used for classification and analysis. PSO has the ability to effectively solve highly non-linear mixed integer optimization problems are typically of complex engineering systems. This model architecture contains input vectors that represent historical data and technical indicators. LS-SVM optimized with PSO is the best one with the lowest error value followed by LS-SVM algorithm. The algorithm combines Particle swarm optimization (PSO) and least square support vector machine (LS-SVM). The PSO algorithm is implies to optimize LS-SVM to predict the daily stock prices. The particle swarm optimization varies from evolutionary computing is they provide flying extreme solutions using hyperspace and makes toward better results, where in evolutionary computation operate directly on solutions which are locations in hyperspace.

V. SYSTEM ARCHITECTURE

User enters the company name for analysing the share details of the respective company. Then the validation process takes place. After validating process, the company's raw data will be fetched by searching through the yahoo

browser, then the collected data will be pre-processed and required data will be displayed to the user. From National Stock Exchange(NSE) stock portal extraction of stock dataset for all the companies takes place , from which the user's specified stock dataset will be selected and processed by algorithm know as LS- SVM and PSO. By using LS- SVM the graphical representation of the respective dataset has been created, which leads to the optimization of the dataset and also solves the non-linear problem. The output of the LS- SVM passes to the PSO, This algorithm work along with LS-SVM to get the better solution from the optimized dataset, eradicates mixed integer problem, unwanted noise and generates suitable patterns for the user specified dataset[6]. The optimized dataset enters the time series analysis and the dataset will be assigned with the respective time period of stock portal dataset. Finally the expected optimal prediction value will be displayed to the user.
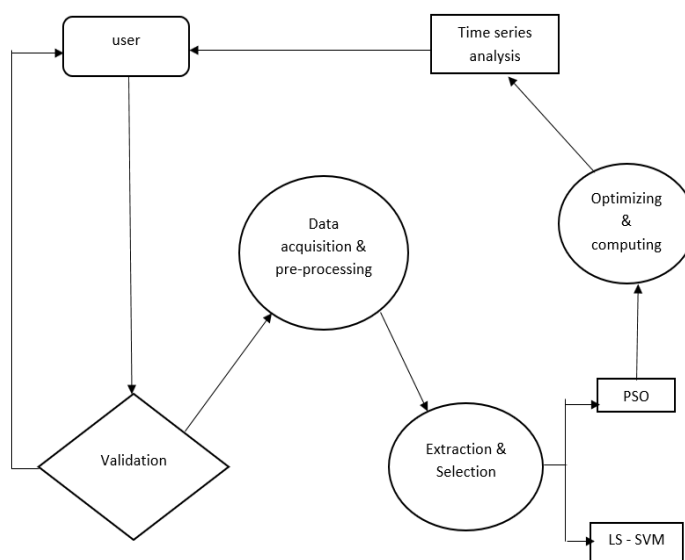

Fig. 1: Architecture of Stock Market

VI. METHODOLOGY

A. Least Square – Support Vector Machine(LS-SVM)

LS- SVM is a supervised learning methods that analyse some information and recognize patterns, and it is used for classification and regression analysis. In this version instead of a convex quadratic programming (QP) problem for classical SVMs, we can use set of linear equations for finding the solutions. Least squares SVM classifiers, LS-SVM algorithm calculate the function from a similar minimization problem found in the SVM method . However the main difference is that instead of using inequalities we can use LS-SVM which involves equality constraints, and it is based on a least square cost function. Moreover, the LS-SVM method solves a linear problem while conventional SVM solves a quadratic one.

the equality constraints of LS-SVM are defined as follows:

$$\text{Min } j\ (w,e,b) = \tfrac{1}{2}\ w^T w + C\ \tfrac{1}{2}\ e^T$$

C∈ℝ is the tradeoff parameter between thesolution size and training errors. From a Lagranian isformed, and differentiating with respect to w,b,e,a (a isLargrangian multipliers).

B. Particle Swarm Optimization(PSO)

PSO is a relatively recent heuristic search method which is derived from the behaviour of social groups.In a single iteration PSO moves from one points to another points with likely improvement using a combination of deterministic and probabilistic rules. The PSO has been popular in industry, mainly because of implementation, and the

ability to effectively solve highly nonlinear, mixed integer optimization problems that are typical of complex engineering systems.it is evolutionary algorithm.Each individual (called particle) is characterized by itspositionX_i, its velocity V_i, its personal best position P_i and itsneighbourhood best position P_g.

$$\upsilon_{ij} \leftarrow \omega\upsilon_{ij} + c_1 q(x_{ij}^{pb} - x_{ij}) + c_2 \gamma(x_j^{sb} - x_{ij}), j = 1,.., n$$

Where w is the inertia weight, X_i^{pb} is the best variable vectorencountered so far by particle i , andX^{sb} is the swarm bestvector, i.e. the best variable vector found by any particle inthe swarm, so far C_1 and C_2 are constants, and q and r arerandom numbers in the range (0, 1). Once the velocities havebeen updated, the variable vector of particle i is modified according to:

$$x_{ij} \leftarrow x_{ij} + \upsilon_{ij}, j = 1,\ldots, n.$$

The cycle of evaluation followed by updates of velocitiesand positions (and possible update of X_I^{pb} and X^{sb}) is thenrepeated until a satisfactory solution has been found.

C. Time Series Analysis

Any metric that is measured over regular time intervals forms a time series. Analysis of time series is commercially importance because of industrial need and relevance especially with respect the demand, sales, supply etc).Each data point (*Yt*) at time *t* in a Time Series can be expressed as either a sum or a product of 3 components, namely, Seasonality (St), Trend (Tt) and Error (et).

For Additive Time Series,

$Y_t = S_t + T_t + \epsilon_t$

For Multiplicative Time Series,

$Y_t = S_t \times T_t \times \epsilon_t$

VII. MODULE DESCRIPTION

A.DATA ACQUISITION & PRE-PROCESSING

1.Data acquisition is a process of collecting data from the yahoo financial website and analysis of the raw data. Then the input data is given for the pre-processing.

2. Pre-Processing plays a key role in overall process .Here all the data are processed simultaneously to remove the unwanted or noise in the data. This process enhances the quality of the data.

B.EXTRACTION & SELECTION

Pre-processed data then given to the extraction in which the data are processed by algorithm such LS-SVM (Least Square – Support Vector Machine) & PSO(Particle Swarm Optimization). By these algorithms ,we can get the meaning full information about stock by making efficient process and pattern organization. The meaning full information are gathered by the maximizing the desired data & minimizing the undesired data. This module plays a major role in the stock prediction.

C.OPTIMIZING & TRAINING

Optimization is a process of getting most cost effective value or information from the extracted value which is high performance solution. This process is used for solving the business problem in the stock market. The Non-linearity problem also can be solved in this process. The optimized values are trained by the training algorithm to face the system interface and some other problem which are related to the stock market.
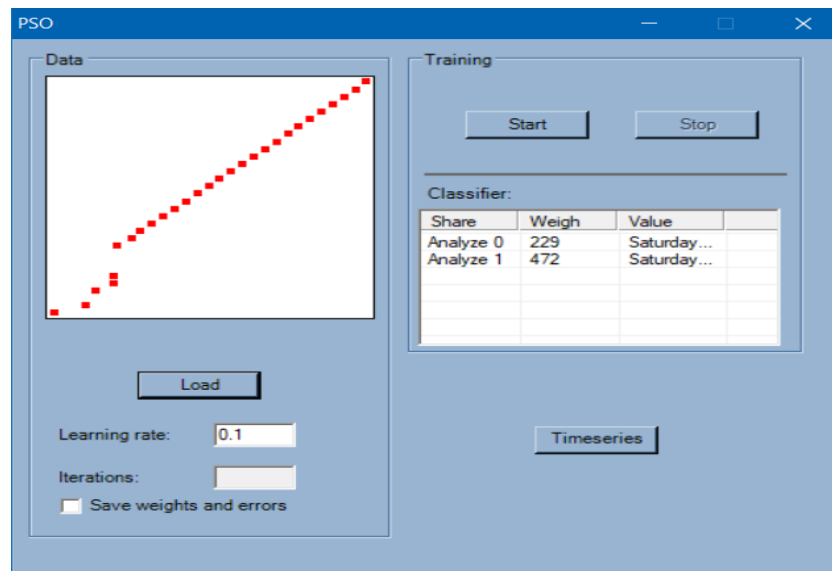
Fig. 2 : Optimizing & Training form

D.COMPUTING & VALIDATION

The information obtained from the previous stages has to be given to the computing process. Time Series Analysis algorithm is used here for validating the information.This algorithm take any of the time interval data and match it with the optimized data for making the data more effective, optimal and accurate. This algorithm represent the information in the form of the graph structure. After validating the information, the optimal information is shown to the client who can invest or sell his share.
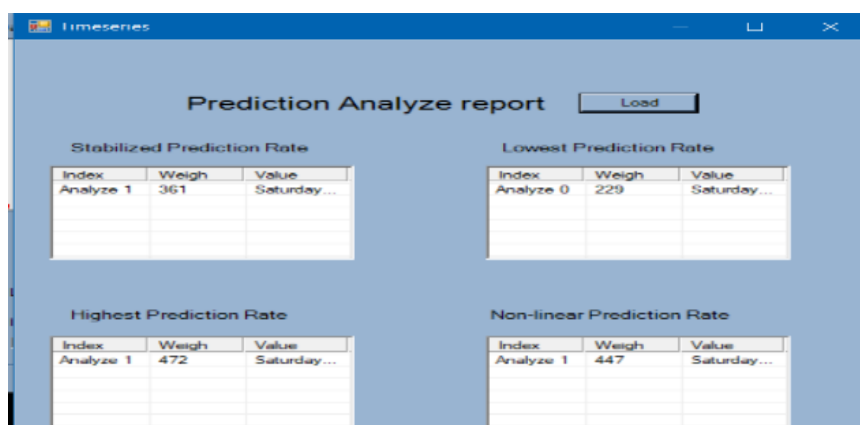

Fig. 3 : computing &validation

VIII. CONCLUSION

In this paper, a generic model for stock market price prediction is offered. Thus using the live date from the yahoo financial feed and the datasets obtained from the NSE used alongside with the LS-SVM PSO for the stock price prediction which is an advanced method, provides the results flexible, feasible, optimal, consistent and error free. Using this system the stock price prediction achieved at faster rate. Proposed system is more trust worthy then stock broker. The accuracy rate of the system is around 90.5 – 93%.

REFERENCES

1. Shilpa Amit Verma Thadomal ,G. T. ThampiThadomal , Madhuri Rao, "Inter-comparison of Artificial Neural Network Algorithms for Time Series Forecasting: Predicting Indian Financial Markets", International Journal of Computer Applications (0975 – 8887) Volume 162 – No 2, March 2017.
2. M.SureshBabu, Dr. N.Geethanjali,Prof B.Satyanarayana(2012),"ClusteringApproachtoStock Market Prediction", Int. J. Advanced Networking and Applications,Volume:03,Issue:04,Pages:1281-1291.
3. Mehak Usmani; Syed Hasan Adil; Kamran Raza; Syed Saad Azhar Ali "Stock Market Prediction Using Machine Learning Techniques" 2016 3rdInternational Conference on Computer and Information Sciences (ICCOINS) Pages: 322 - 327, DOI:10.1109/IC-COINS.2016.7783235.
4. Wanjawa, Barack Wamkaya(2012), "A Neural Network Model for Predicting Stock Market Prices at the Nairobi Securities Exchange", University ofNairobi research archive, Volume – 73404, Page-56.
5. Mahajan Shubhrata D., Deshmukh Kaveri V., Thite Pranit R., Samel Bhavana Y., Prof. Chate P.J(2016), "Stock Market Prediction and Analysis Using Naive Bayes", International Journal on Recent and Innovation Trends in Computing and Communication, Volume – 4 Issue – 11, Issue Number - Issue: 11121-124.
6. Cherkassky, V. and Ma, Y., "Practical Selection of SVM Parameters and Noise Estimation for SVM regression". Neural Networks, vol., 17, pp. 113-126, 2004.