



An enabling technique analysis in Data Mining for Stock Market trend by Approaching Genetic Algorithm

M.Tawarish¹, Dr. K. Satyanarayana²

¹Research Scholar, Department of Computer Applications (Ph.D.), Bharath Institute of Higher Education and Research (BIHER), (Declared as Deemed-to-be-University under Section 3 of UGC Act 1956)
173 Agharam Road, Selaiyur, Chennai - 600 073, Tamil Nadu, India.
mtawarishir@yahoo.co.in

²Research Supervisor / Principal, Sindhi College of Arts and Science,
Chennai – 600 077, Tamil Nadu, India.
ks62@rediffmail.com

ABSTRACT

A trend is the common direction at which the stock is moving. Based on whether the market is optimistic or pessimistic, the trends move upwards or downwards. Share market trend analysis is a part of technical analysis that tries to predict the future development of a stock based on past data. This study try to help the investors in the stock market to choose the better timing for buying or selling stocks based on the knowledge extracted from the chronicled prices of such stocks. The decision taken will be focused on decision tree classifiers which are one of the data mining techniques and by impending genetic algorithm.

Key words: Data Mining, Data Mining techniques, Decision tree classifier, Genetic algorithm, Share market trend.

1. INTRODUCTION

Highlight The stock market refers to public markets that exist for issue, buying and selling stocks that trade on a stock exchange or over-the-counter. Stocks, commonly known as equities, represent fractional ownership in a company, and the stock market is a place where investors can purchase and trade ownership of such investible assets.

An efficiently running stock market is considered vital to economic development, as it gives companies the ability to quickly access capital from the public.

The stock market trend [1] is what technical analysis being about. As a practical investor, you would look at chronological price patterns and form an opinion about market trends. Accordingly, you will decide your route of action about a stock. This is why considerate stock market trends is so significant.

What this section will be about—what are market trends; application in stock selection; how to identify market trends, and how to examine them to make profitable investments. Share market trend analysis is a phase of technical analysis that tries to predict the anticipate progress of a stock based on precedent data. A share market trend is based on the concept that the past movements are skylight to the future trends

1.1 Impacts on the Stock Market

In some factors if we take decisions of buyers and investors about where to buy and sell individual stocks. The most important are the profits of the company, and/or its predictions for profits in the future. Fluctuations in “the market” in a more common sense are simply the sum of all those individual stock decisions, but there are things that edict the direction of the market in general. To understand them the market is a forward-discounting mechanism.

In lay person terms, buyers are always looking forwards. They use previous price action in the form of charts to inform their decisions to sell or buy, but what decides whether they make the exact decision or not is what will happen in the forthcoming.

Apparently, the future is unknown, but some things are used to determine the possibility of a stock's price moving forward or downward. Many of individual's things, such as economic strength and political stability affect all stocks, so changes in things like economic conditions can cause widespread buying or selling, and thus general market moves.

2. DATA MINING

Data mining is that observe of mechanically looking out massive stores of information to find patterns and trends that

transcend straightforward analysis. Data processing uses refined mathematical algorithms to section the information and assess the chance of future events.

Data mining is commonly known as Knowledge Discovery Database.

The main properties of Data Mining [2] are:

Automatic Detection: Data mining is trained by structured models. A model uses an algorithm to follow up on an arrangement of information. The idea of automatic detection refers to the execution of data mining models. Data mining models can be used to mine the data on which they are constructing, but most types of models are simplifying to new data.

Prediction: is an area of facts that deals with extract information from data and using it to predict trends and performance patterns. Often the strange event of interest is in the future, but predictive logic can be applied to any type of obscure facts whether it be in the past, present or future.

Actionable info: Data mining will derive unjust data from massive volumes of information. For instance, a city planner would possibly use a model that predicts financial gain supported by demographics to develop an idea for low-income housing. An automobile leasing agency may use a model that identifies client segments to style a promotion targeting high-value customers.

3. GENETIC ALGORITHM

Genetic Algorithm (GA) [3] is a search-based optimization technique based on the principles of Genetics and Natural Selection. It is normally used to find best or near-optimal solutions to difficult problems which otherwise would take a time to solve. It is often used to solve optimization problems in research area and machine learning sources. Optimization, in any process, we have a set of inputs and a set of output.

Optimization refers to finding the values of inputs in such the way that we tend to get the “best” output values. The definition of “best” varies from drawback to drawback, however in mathematical terms, it refers to maximizing or minimizing one or a lot of objective functions, by varied the input parameters. The set of all potential solutions or values that the inputs will take form up the search area. During this search area, lies some extent or a collection of points which provides the best answer. The aim of improvement is to search out that time or set of points within the search area.

3.1 Merits and Limitations in Genetic Algorithm

Faster and a lot of economical as compared to the normal strategies. Optimizes each continuous and distinct functions and additionally multi-objective issues Provides a listing of “good” answers and not simply one solution. Always gets a solution to the matter that gets higher over the time. Helpful once the search area is incredibly massive and there are an oversized variety of parameters concerned. Genetic Algorithms aren't suited to all issues, particularly issues that are easy and that spinoff data is offered. Being random, there aren't any guarantees on the optimality or the standard of the answer. If not enforced properly, the genetic algorithm might not converge to the best answer.

4. LITERATURE SURVEY

This research work presents a data mining based stock market trend prediction system, which produces highly accurate stock market forecasts. The future system is a genetic algorithm (GA) optimized decision tree-support vector machine (SVM) hybrid, which can predict at some point ahead trends in stock markets. The uniqueness of the future system lies in the use of the fusion system [4] which can adapt itself to the changing market conditions and in the fact that while most of the attempt at stock market trend prediction have approached it as a regression problem, present study converts the trend prediction task into a classification problem, thus humanizing the prediction accuracy considerably. Technical indicator is used in the present study to extract data.

The research deals with the design and validation of a data mining decision system which generates buying and selling orders in the stock market. The decision system's design has three main phases as: the data processing phase, the data mining phase and the evaluation phase. The data processing phase selects from the raw data base a dataset that focuses on a subset of attributes or variables on which knowledge discovery is to be performed. It also removes outliers and redundant information, and uses financial indicators to represent the processed data by means of an Object Attribute Table (OAT). The data mining phase converts the data contained in the OAT into useful patterns; in particular decision trees are found [5]. The assessment phase proves the consistency of outline by means of a testing set. The positively evaluated decision system can then be used in real world situations that will allow for its validation.

The research by used decision tree technique to build on the work of Lin where Lin tried to modify the filter rule that is to buy when the stock price rises k% above its past local low and sell when it falls k% from its past local high showed Lin's method out- performed the filter rule. According to, in Lin's

work [6], the criteria for clustering trading points involved only the past information; the future information was not considered at all. The research by aimed to improve the filter rule and Lin's study by considering both the past and the future information in clustering the trading points.

Kannan, Sekar, Sathik and P.Arumugam in used data mining technology to discover the hidden patterns from the historic data that have probable predictive capability in their investment decisions. The forecast of stock market is testing task of financial time cycle predictions. At present five methods are Typical price (TP), Bollinger bands, Relative strength index (RSI), CMI and MA used to analyze the stock index. In this article the author got the profitable signal is 84.24% using Bollinger Bands rather than MA, RSI and CMI [7]. Prediction stock price or financial markets has been one of the biggest challenges to the community.

A variety of technical, basic, and statistical indicators have been proposed and used with unstable results. Surveyed some recent literature in the domain of machine learning techniques and Genetic Algorithms used to predict stock market movements. Data mining are identified to be the dominant machine learning technique in stock market prediction area.

5. METHODOLOGY USED

The trend prediction system methodology used in this paper works in the way: First the modules are mined from the daily stock market data. Then the related features are selected by using decision tree. A rough set classifier is then used to predict the next day's trend using the chosen features. Then these trends are evaluated using data mining classifications, impending in GA and final predictions will generate. In the current study, the prediction exactness of the future system is validated using the Bombay Stock Exchange Sensitive Index (BSE-SENSEX or commonly, SENSEX) data.

The performances of trend prediction systems are valued using the cross validation method [8].

1. Selection Data
2. Data Mining Prediction
3. Extraction
4. Evaluation



Figure 1: BSE SENSEX Chart

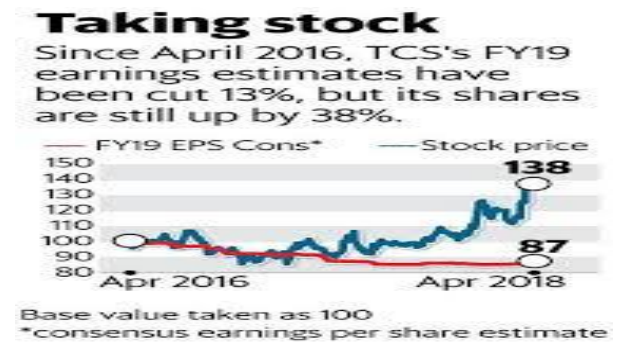


Figure 2: TCS SENSEX Chart



Figure 3: ITC SENSEX

5.1 Data Studies

Selection Stock market data for BSE, TCS, ITC as shown in above Figures.

The above all the data sets will be extracted and transformed to the prediction of data mining classifiers to obtain a result by impending Genetic Algorithms

5.2 Time Series Forecasting – market trend analysis

Time Series Forecasting (TSF) Figure 4 is a process of analyzing time series data and predicting the future result. Using of statistical techniques to model and clarify time-dependent series of data points. Time Series Forecasting is a widely used technique in market forecasting.



Figure 4: Trend moving upward and downward

Time Series is a series of sharp data points measured at constant time interval over a period of time. Time series analysis is the using of statistical techniques to study time series data and extract significant statistics and time about the data.

5.3 Evaluation results of TSF

There were three methods used for estimate of forecasting accuracy in this study, are follows

(a) The Formula for Mean Absolute Error is given as,

$$MAE : \frac{1}{n} \sum_{t=1}^n |O^t - y^t|$$

(b) The Formula for Root Mean Squared Error is given as

$$RMSE : \sqrt{(P_i - A_i)^2}$$

P is Predicted value
A is Actual value

(c) Direction Accuracy

Direction Accuracy is by this principle arithmetic predication as

$$= \text{Counttot}(\text{sign}(y_i - y_j) == (\text{sign}(y_i - y_{i-1}))) / n$$

\hat{y} = predicted value of i

5.4 Decision Tree Rough set based classifier [9]

The methodology used here by, mathematical rough set based classifier is applied to calculate future values of stock.

Method is worked as follows

- (a) Calculate the number of items in data set.
- (b) Cross Validation by

5.5 Regression analysis calculation:

Regression analysis can be used to model the relationship between one or more independent or predictor variables and a dependent or response variable.

Let D be a training set consisting values of predictor variables, set contain P data points of the form (x1,y1), (x2,y2), ..., (x|P|, y|P|).

The regression coefficient can be estimated by the following equation in the Figure 5 and 6.

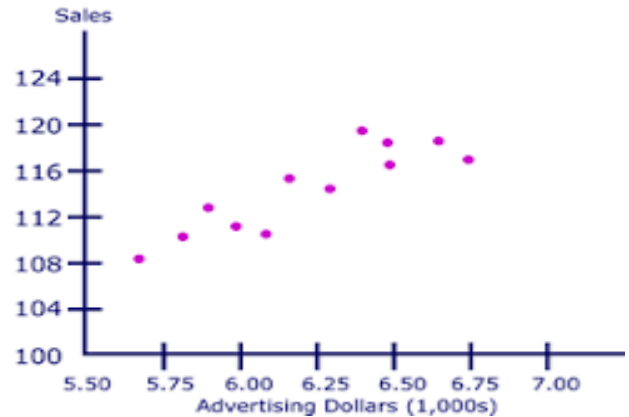


Figure 5: Regression Linear Chart

$$Y = a + bX$$

$$Y = (\bar{Y} - b\bar{X}) + bX$$

$$Y = \bar{Y} + b(X - \bar{X})$$

$$Y = \bar{Y} + bx$$

Figure 6: Regression equation

Linear regression equation Figure 7 is one of the mainly regular data mining techniques for predicting the upcoming value of variable based on the linear connection it has with other variable.

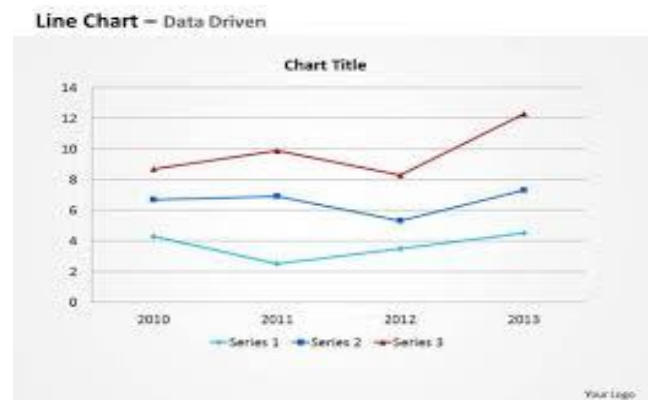


Figure 7: LR Method

Correlation Validation Formula and the graph represented in Figure 8 and Figure 9.

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Figure 7: Correlation Equation

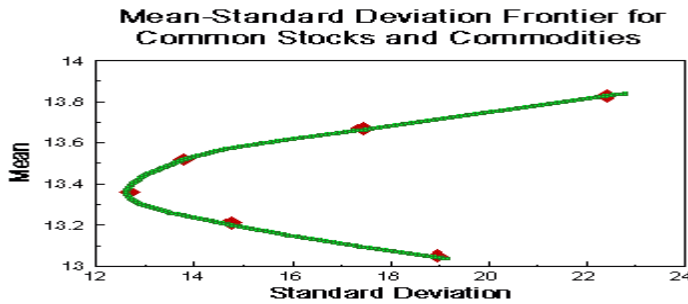


Figure 9: Graph Correlation

Where x and y are the sample means for the sample standard deviation for X and Y.

5.6 CART - Decision Tree classifier

5.6.1 Decision Trees:

Classification and Regression Trees or CART [10] for is a term introduced by Leo Breiman to refer to Decision Tree algorithms that can be used for classification or regression predictive modeling problems. CART decision tree methodology are commonly used in data mining with the objective of creating a model that predicts the value of a target or dependent variable based on the values of several input (or independent variables). CART decision tree methodology.

5.6.2 Classification Trees:

Classification tree methods are suggested when the data mining task contains classifications or predictions of outcomes, and the goal is to generate rules easily. A Classification tree is built during a process known as binary recursive partitioning. This is an iterative process of split the data into partition, and then splitting it up further on each of the twigs Figure 10.

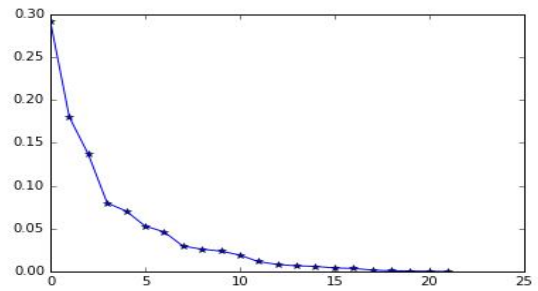


Figure 10: Classification Tree

5.6.3 Regression Trees [11]:

A regression tree is built through a process known as binary recursive partitioning, which is an iterative process that splits the data into partitions or branches, and then continues splitting each partition into smaller groups as the method moves up each twig. The Regression Tree Algorithm can be used to find one form that results in good predictions for the new data. We can view the current predictor Figure 11 to see if our model is a good fit to the data.

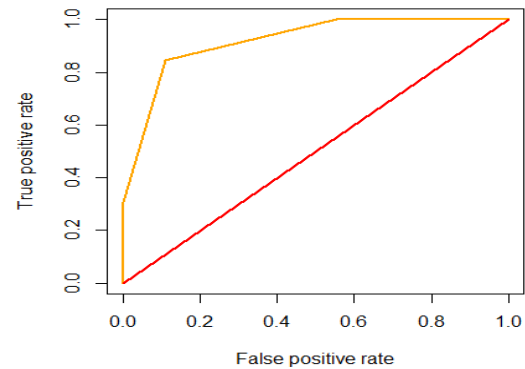


Figure 11: Regression Tree Predictor Fit

6. RESULTS

The Data set studies results are predicted for stock market trend by using the model through evaluation of TSF (market trend analysis) which gave the approximate results (Table 1), trend to move upward or downward by MAE, RMSE and Direction Accuracy.

Table 1: TSF Results

Class	Number of Objects	Lower Trend (approx.)	Upper Trend (approx.)
Up	752	712	773
Down	190	188	198
No Trend	700	687	707

The Evaluation result on set of test for BSE (Stock Exchange) in Table 2

Table 2: BSE evaluation result

Mean absolute error	0.995
Root mean square error	27.08
Direction Accuracy(DAC)	3.55

6.1 PREDICTED FOR ROUGH BASED CLASSIFIER

By giving sample input BSESENSEX data Table 3 taken for the study. Data set contains five attributes, namely serial number, year, price (sp), earnings and dividend. Three features earning. Price and dividend is use for prediction process. Dataset contains values from year 1979to 2018.Dataset contains 39 instance.

Table 3: BSE Sensex Table

Transactions	Year	Stock price	earning	dividend
1	1979	45987	23000	11887
11	1980	587254	25000	13765
12	1981	1458756	31000	23454
14	1982	56898	430000	12345
12	1983	850007	234567	32000
6	1984	458789	235667	34566
7	1985	1458789	12456	10909
5	1986	544856	456778	23555
9	1987	555456	32345	18790
2	1988	489332	23456	23417
11	1999	478965	245678	234568
14	2000	698789	2345678	123543
15	2001	145621	76545	45678

By Regression analysis and correlation validating for predicting the upcoming value of variable based on the linear connection it has with other variable.

The Graph chart for the Classification and regression tress

are shown in the Figure 13. It represents for TCS and ITCL and other companies comparing data set moving.

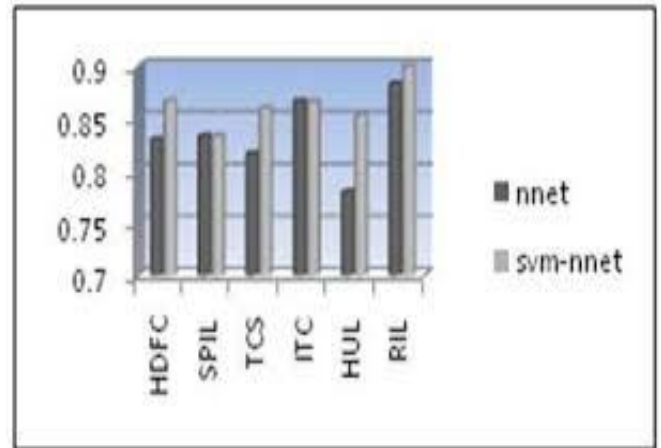


Figure 13: Cart Data TCS , ITCL compared with other Limited's

The Output for CART and Rough Set Classifier as shown in the Table 4: Transactions Worked.

Table 4: Rough Set Classifier Table

Transactions	Year	Stock price	earning	dividend
11	1980	587254	25000	13765
12	1981	1458756	31000	23454
14	1982	56898	430000	12345
12	1983	850007	234567	32000
6	1984	458789	235667	34566
7	1985	1458789	12456	10909
5	1986	544856	456778	23555
9	1987	555456	32345	18790
2	1988	489332	23456	23417
11	1999	478965	245678	234568
14	2000	698789	2345678	123543
15	2001	145621	76545	45678
16	2002	14587965	34587	21765

7. CONCLUSION

This research paper critically studies the major determinants of stock market performances in India which has been in a fluctuating trend in the recent years. Stock markets seem to be highly complex and unpredictable. So we have studied various methods for stock market prediction using Data Mining and genetic Algorithm. Because of usefulness and needs from the people, opinion mining techniques became an active research area. As the size of the following data increases, analyzing and summarizing opinionated data is becoming more vital. To satisfy these needs, various kinds of techniques are proposed.

The approaches of using TSF, mining classifier, statistics techniques of correlation and regression analysis, finally by CART Methods we have been predicted. But, we have to conclude that stock prediction is very difficult task and various influences should be considered for forecasting the market more accurately and efficiently. In future work, we intend to determine the vital impact of particular basic analysis variables on quality of stock price prediction.

REFERENCES

1. Qasema, Al-Radaideh, Adel Abu Assag, Eman al nagi. The International Arab conference on IT “**Predicating stock prices using data mining techniques**”,
2. <http://data-mining.philippe-fournier-viger.com/introduction-data-mining/>
3. Li Lin, Longbing Cao, Jiaqi Wang, Chengqi Zhang “**The Applications of Genetic Algorithms in Stock Market Data Mining Optimisation**” Faculty of Information Technology, University of Technology, Sydney, NSW 2007, Australia Capital Market CRC, Sydney NSW 2000, Australia.
4. Binoy B. Nair 1 V.P Mohandas 2 N.R. Sakthivel “**A Genetic Algorithm Optimized Decision Tree SVM based Stock Market Trend Prediction System**” Amrita School of Engineering, Amrita Vishwa Vidyapeetham Coimbatore, Tamil Nadu, India.
5. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 12, 2011147 | P a g e www.ijacsa.thesai.org “**Detection and Extraction of Videos using Decision Trees**”.
6. Wu, M.C., Lin, S.Y., and Lin, C.H., (2006) “**An effective application of decision tree to stock trading**”, Expert Systems with Applications, 31, pp. 270-277.
<https://doi.org/10.1016/j.eswa.2005.09.026>
7. “**Financial Stock Market Forecast using Data Mining Techniques**” K. Senthamarai Kannan, P. Sailapathi Sekar, M.Mohamed Sathik and P. Arumugam.
8. **Prediction of the Bombay Stock Exchange (BSE) Market Returns Using Artificial Neural Network and Genetic Algorithm** Yusuf Perwej1, Asif Perwej2.
9. <https://www.solver.com/classification-tree>
10. **Using CART for Stock Market Forecasting February 28, 2014** By The R Trader
11. <https://www.stat.cmu.edu/~cshalizi/350/lectures/2/lecture-22.pdf>