

A Review on Pricing Prediction on Stock Market by Different Techniques in the Field of Data Mining and 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

In the current developing focused market, forecasting or predicting the stock returns just as the organization's economic status will give more profits for the investors to contribute positively. Stock forecast can be done by utilizing the present and past information accessible available. Forecasting the stocks should be more exactness, scalability and less time utilization. Various enabling techniques has been taken in the field of data mining for predicting the stock market by approaching the genetic algorithms. Such as Decision Tree Classifiers, Artificial neural networks-ARIMA, Bollinger bands, RSI and so on. This survey paper review helps to know the best enabling technique tools for the predicting the stock mining how to invest, when to invest for the investors in the future. Many of the researchers has been analyzing the predictions of stock market by getting the help of Mining through genetically.

Key words: Data Mining, Enabling Techniques, Genetic Algorithm, Review, Stock Market.

1. INTRODUCTION

Predicting the stock market is an essential field of sole, the longer term price of the value for higher stock of finance between organizations having the shares. Within the field of prediction, the financial standing, stock market prediction has become a significant one. Once the financial status of a corporation has been accurately foreseen, the investors are going to be confidently investing their currency, although several tries are created to predict the economic market's performance, none of the techniques have provided a sure result since it is a tough task due to screaming and time variable knowledge and additionally every technique used has its own demerits. The stock market prediction is formed by two groups. 1. Trend Forecasting model: This model is for forecasting the link among different techniques

variables and the stock price enhancement. 2. Time series Prediction: It predicts by examine the actual turning of stock and used for predicting the future stock price return [1].

A various common models and techniques are used for predicting or forecasting the stock which includes Decision Tree Classifiers in Data Mining, Artificial Neural networks, Genetic Algorithms and Bayesian Networks. Recently, it's been found that a productive choice of options related to predicting stock growth can facilitate providing other accurate result. in this perspective, several programs regulation has also been utilized in order to decide on actual features which has a lot of impact on the movement of stock returns.

Many researchers have predicted the stock trend by hybridizing the feature of ARIMA, CART, Bollinger Bands and RSI models in which to provide accuracy to a certain extent in this review paper.

2. METHODOLOGIES - STOCK MARKET PREDICTION

Various Techniques used for the prediction of stock market and the methods are made clear in the Table 1.

Table 1: Various Techniques used for predicting stock market

S.No	Reference	Dataset	Methods	Performance
1	R 2	TSF	TM,RSE,DA	Good Performance
2	R 3	BSE	CART, RBC	App. Trend
3	R 4	CPCL	CHI-Compound	Avg 65%
4	R 5	AGPSISA	ARIMA,RMSE	Perfect
5	R 6	HDFC	TP, Bollinger	Stable
6	R 7	HDFC	RSI MA	80% Good

3. RELEATED WORK - Techniques

3.1 Data Mining Process and its Application Techniques

There are three stages for data mining process. (a) Initial Analysis (b) Typical Building (c) Lineup. These applications of models are new data to generate predictions [8].

3.1.1 Artificial neural networks

These are natural systems that detect patterns, make predictions and study. The artificial notably are computer programs implementing difficult pattern detection and machine learning algorithms on a PC to build predictive techniques from huge historical data bases. Neural networks grow out of the area of Artificial Intelligence rather than from the control of statistics. These are Nonlinear predictive techniques that learn through training and simulate natural neural networks in System.

3.1.2 Decision trees

A decision tree is an analytical model that can be view as a tree. Tree shaped models represent a set of decisions. These decisions create rules for the arrangement of a data set. Explicit decision tree methods contain Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID). Linear regression and Correlation is a statistical models used to find the best-fitting linear relationship between a target variable and its predictors.

3.1.3 Applications

A large range of groups have expanded successful applications of data mining. The applications of data mining include: banking industry, credit card Association, financial trade, Automation industries, moving business, turnover management, fraud detection, parcel Goods Company, government policy setting, hiring profile, medical administration, process control, Quality control, store organization, student recruiting and retention, guarantee analysis and so on.

3.2 Genetic Algorithm

Inflation techniques that utilize process such as genetic permutation, mutation and natural selection in a plan based on the concepts of development. The generative techniques used for this research are the genetic algorithms and evolution strategy. By means of these algorithms, we are annoying to find the connection credence for each point, which helps us in predicting the final aim price of the stock. An input for each quality is given to a sigmoid function after it is amplified based on its connection weight. The investigational results show that using this genetic approach, predicting the stock price is hopeful.

Genetic Algorithm [9] predicts the stock market price to a certain level with the help of Mining Databases. Genetic Algorithms works genetic programming by(a) Primitive (b) Fitness Functions (c) Initial Populations (d) Evolve Peoples (3) Terminate condition. They modeled the system with basic Genetic Algorithm by four stages such as initialization, selection, Crossover, and mutation for the SENSEX data of the Companies used in prediction of stock markets.

3.3 CART/Rough Based Classifier

3.3.1 Classification and Regression Trees

Classification and Regression Trees [10] is a set of module for which classification and prediction technique. This technique is intended to produce conventions that predict the cost of a result (target) variable from identified values of predictor variables. There are many special applications but they are all sharing a general characteristic usage. An approach based on classification and regression trees shows several profits if compared with the parametric and nonparametric approach that have already been introduced. As shown by Breimanetal. (1984), who introduced the method, classification trees can be used with any data set given a satisfactory vector of questions.

This means that the model can be applied to both definite data and structured data, which in its turn has the outcome that one does not have to be clear to a certain grouping when conducting the study, the another one model is very fine in handling non identical relationships with respect to provisional information. Different algorithms use different ways for measuring “elite”. These generally measure the consistency of the target variable within the divisions. These terms are applied to each candidate division, and the resulting values are combined as averaged to provide a measure of the quality of the rift.

3.3.2 Rough Based Classifier

Rough set is a latest mathematical tool for dealing with abstraction and uncertainty predictions. The study involves for the use of provide decision hold for stock dealers, analyzes the economic data of stocks with the rough set theory and class the data according to their impact for the market performance of stocks. The important responsibilities include, creation of the stock trading data set, pre-processing of stock dealing data, analysis of the economic data set with the data mining software set, analytical results and confidence, and conclusions. The investigational results and experimental results of the study indicate that this study provides an easy way for users of the securities market in predicting stocks.

A rough set drop technique to find all cutbacks of the data that contain the least subset of attribute linked with a class label for classification. To evaluate the legality of the rules based on the

estimate quality of the attributes, we set up a statistical test to evaluate the significance of the system. An uncertain result from applying the rough set approach to the set of data samples are given and evaluated. In additionally, the rough set classification accuracy is also compared to the well-known significant techniques. The system of study exhibited that the theory of rough sets is a useful tool for inductive purpose of predict and a valuable aid for predicting the stock market prices of the data set.

3.4 ARIMA TECHNIQUES

ARIMA models [11] are, in generally the most common class of models for forecasting a time series in stock market trend which can be made to be “static” by changing perhaps in combination with nonlinear transformations such as listing or collapsing. A random variable that is a time series is static if its statistical properties are all constant eventually. A static series has no trend, its variations around its mean have a constant amount, and it twists in a consistent method, that is its short-range random time displays always look the same in an arithmetical sense. The final condition means that its autocorrelations remain constant eventually, or equivalently, that its switch scale remains continuous over time. A random variable of this form can be viewed as a combination of signal and sound, and the signal could be a pattern of fast or slow mean return, or overall fluctuation, or fast alternation in sign, and it could also have a cyclic component. An ARIMA model can be viewed as a “clean” that tries to separate the signal from the whole, and the signal is then inferred into the future to obtain forecasts. The ARIMA forecasting equation for a stationary time series is a linear regression equation in which the predictors consist of intervals of the dependent variable or intervals of the forecast errors. (i.e.) Predicted value of $Y = a$ constant and weighted sum of one or more recent values of Y or a weighted sum of one or more recent values of the errors.

If the predictors consist only of protected values of Y , it is a pure autoregressive self-regressed classic, which is just a special case of a regression model and which could be fitted with standard regression software. For example, a first-order autoregressive (ARIMA) model for Y is a simple regression model in which the independent variable is just Y wrapped by one period ($LAG(Y,1)$ in Stat graphics or $Y_INTERVAL1$ in Regress It). If some of the predictors are lags of the errors, an ARIMA Technique it is NOT a linear regression model, because there is no way to specify last time's error as an independent variable: the errors must be computed on a time level basis when the model is fixed to the data. From a technical viewpoint, the drawback with using protected errors as predictors is that the model's predictions are not direct functions of the coefficients, even though they are linear functions of the past data. So, coefficients in ARIMA models that include wrapped errors must be estimated by real valued functional techniques somewhat than by just explaining a system of equations.

The abbreviation ARIMA stands for Auto-Regressive Integrated Moving Average. Intervals of the COLUMN series in the forecasting equivalence are called "autoregressive" terms, lags of the forecast errors are called moving average expressions, and a time series which needs to be differenced to be made column is said to be an integrated version of a stationary series.

3.5 Bollinger Band/RSI Techniques

3.5.1 Bollinger Bands

Bollinger Bands [12] a technical pointer. They predict future market actions exclusively based on past market data, which they collective and calculate. The result is drawn straightforwardly into your price chart.

The pointer creates three lines, which are normally indicated to as “bands” “Middle line. This is the foremost streak. It is a moving average, usually based on twenty intervals. Upper line. This streak is the result of the moving average and the standard deviation enlarged by an issue, normally 2. Lower line. This streak is the result of the moving average loss the standard deviation multiplied by an issue, usually 2.

Bollinger Bands can be a great help for dualistic decisions dealers. They can Locate new trading openings. When the market methods a Bollinger band, you know that the market is likely to seizure around. This information alone is enough to win a dualistic decision. Strategy curve. Bollinger bands require a clear indication for how far the markets can effort. Dualistic options types with high outgoings such as hierarchy options or one touch selections require this prediction, which is Bollinger Bands can turn a normal strategy into a highly gainful one. Prevent bad trades. Bollinger bands form important conflict and care levels. Even if you choose to not openly trade Bollinger Bands. Then, you might be surprised when the market does something unforeseen, and you lose a trade you supposed to be an assured success.

3.5.2 RSI

The relative strength index [13] is a momentum indicator that measures the level of current price changes to assess overbought or oversold conditions in the price of a stock or other quality. The RSI is exhibited as an oscillator a line graph that moves among two boundaries and can have an analysis from 0 to 100. The indicator was originally developed by J. Welles Wilder and introduced in his seminal 1978 book. The RSI is a popular momentum oscillator developed in 1978. The RSI compares the stock price predictions momentum. Stocks are considered overtrade when the indicator is above 70% and resale when the indicator is below 30%. The average is to use 14 periods to compute the initial RSI value. For instance, see the market closed higher seven out of the earlier 14 days with an average profit of

1%. The persisting seven days all closed lower with an average loss of -0.8%.

4. CONCLUSION

By the utilization of these different accessible techniques, it is conceivable to make another strategy to forecast the future patterns in stock market. It is conceivable to utilize any of the reviewed techniques and progress a hybrid system for the prediction of fiscal status of a firm accurately. Except it is essential to design the system accordingly by which the accurateness and concert can be increased with less computational complexity. This paper is an entire survey of techniques and models used by many researchers in the use of data mining techniques for stock market prediction. This review will also support in selecting the best algorithm for time series and trend prediction technique based on their prediction terms.

REFERENCES

1. A Review on Prediction of Stock Market using Various Methods in the Field of Data Mining. Shobana and A. Umamakeswari*School of Computing, SASTRA University, Tirumalaisamudram, Thanjavur – 613401,
2. Time series forecasting by evolving artificial neural networks with genetic algorithms, differential evolution and estimation of distribution algorithm Juan Peralta Donate • Xiaodong Li • Germa'n Gutie'rrez Sa'nchez • Araceli Sanchis de Miguel
3. <https://pdfs.semanticscholar.org/29a7/e1e1386db50bbb22710d63eb820f9c86ae1b.pdf>
4. <https://in.investing.com/equities/chennai-petroleum-corporation-historical-data>
5. Proceedings of the International MultiConference of Engineers and Computer Scientists 2016 Vol I, IMECS 2016, March 16 - 18, 2016, Hong Kong. Artificial Neural Networks and Time Series Models for Electrical Load Analysis Ronnachai Chuentawat*, Supaporn Bunrit, Chanintorn Ruangudomsakul, Nittaya Kerdprasop, and Kittisak Kerdprasop
6. http://shodhganga.inflibnet.ac.in/bitstream/10603/14788/1/11_chapter%205.pdf
7. <https://www.bollingerbands.com/bollinger-bands>.
8. Financial Stock Market Forecast using Data Mining Techniques by K. Senthamarai Kannan , P. Sailpathi Sekar, M. Mohamed Sathik , Arumugam<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.302.5160>
9. Genetic Programming An Introductory Tutorial and a Survey of Techniques and Applications Riccardo Poli Department of Computing and Electronic Systems University of Essex, UK rpoli@essex.ac.uk
10. <https://www.r-bloggers.com/using-cart-for-stock-market-forecasting/>
11. RIMA models for time series forecasting<https://people.duke.edu/~rnau/411arim.html>.
12. <https://www.math.kth.se/matstat/seminarier/reports/M-exjobb10/100308b.pdf>
13. <https://www.investopedia.com/terms/r/rsi.asp>.