INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Prediction of Stock Market Using Financial News Analysis and Supervised Data Mining Technique

Soniya Dewangan[1], S R Tandan[2], Shagufta Farzana[3], Rohit Miri[4]

[1]*M.Tech Scholar- Department of Computer Science and Engineering*
[2,3,4]*Assistant Professor, Department of Computer Science and Engineering*

Abstract—Financial news articles play very important role in stock market prediction because these financial news affect the decision of investor. Decision making for investor in stock market is considered to be one of the difficult task, because, stock market behaviour is dynamic. The nonlinear nature of the stock market data makes it very difficult to design a system that can predict the future direction of the stock market with sufficient accuracy. Thus we introduce system that used hybrid approach of classification rule of data mining and financial news articles effects. Here we used four years index points of BSE and analyze it for stock price prediction. In this paper we show significant correlation between the changes in daily stock index values and the effect of financial news to the prediction of stock market prices.
Keywords—BSE, financial news, ruled based classification, stock price prediction.

I.   INTRODUCTION

 Data mining can be described as "better use of data". Mining stock market tendency is regarded as a challenging task due to its high volatility and noisy environment.Data mining deals with the discovery of hidden knowledge, unexpected patters and new rules from large database. Prediction of stock market returns is an important issue in finance. Now a day it is one of the measure role for trading the stock markets. Mining news articles and the time series data concurrently, for predicting the stock market prices is an emerging topic in data mining and text mining communities. Basically prediction is accomplish with the help of news regarding day by day changed and company bench mark. Bench mark of company is based its profit during its production and market gross income. These are come under short term investment of the money in the company [1]. One can find these index values for their better prediction and predict it for future use.BSE (Bombay Stock Exchange) is one of the leading Stock exchanges in India. BSE was established in 1975. The BSE has played a leading role as a change agent in transforming the Indian stock market to its present form. The BSE 30 Composite Stock Price Index is an index of 30 stocks from major industries of the Indian economy. The BSE has brought about unparalleled transparency, speed and efficiency, safety and market integrity. In this process the BSE has become the largest stock exchange in the country, relegating the National Stock Exchange to the second place. Here we are taking stock market prices from July 2011 to July 2015 for prediction.

Financial stock market is a complex, non inactive, noisy, disordered, nonlinear and dynamic system but it does not follow random walk process [9, 10]. There are many factors that may cause the rise and fall of financial market movement. Predictions of stock market price and its direction are quite difficult. Data mining techniques have been profitably have to shown to generate high prediction accuracy of stock price movement [12]. Recently some of the researchers have found that news are one the most influential sources that affect stock market and are necessary in achieving to more accurate predictions[21].

The prediction of stock market is without doubt an interesting task. There are number of methods applied to accomplish this task. These methods use various approaches, ranging from highly informal ways to more formal ways. These techniques are:-technical analysis method, fundamental analysis method, Traditional time series methods and machine learning methods [23]. Technical analysis is an alternative approach to the study of stock price behaviour. The rational behind technical analysis is that share price behaviour repeats itself over time and analysis attempt to derive methods to predict this repetition. A technical analyst looks at the past share price data to see if he can establish any patterns, then looks at current price data to see if any of the established patterns are applicable and ,if so, extra potations can be make to predict the future price movements[18]. Technical analysts, known as chartists, attempt to predict the market by tracing patterns that come from the study of charts which describe historic data of the market. Charting represents a key activity in technical analysis, because graphical representation is the very basis of technical analysis. Charting techniques work with the visual representation of the

541

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

stock price graph over a selected period. By enhancing the diagrams with secondary time series documenting perceived trends, the chartists identify trends or trading opportunities.

Data mining refers to the process of finding interesting patterns in data that are not explicitly part of the data [12, 13]. Data mining uses two strategies: supervised and unsupervised learning. In supervised learning, a training set is used to learn model parameters whereas in unsupervised learning no training set is used (e.g., k means clustering is unsupervised) [15].Classification techniques is a type of supervised learning of data mining. Classification is define by the set of predefined classes from a training set as a result of learning from that dataset[15].Financial news articles are one of the key factor to effect the stock market. A successful news analysis would be achieved if the effective information about stock could be extracted from the news article. It is must to extract the key phrases which may effect the stock price because we know that news articles are unstructural or textual information. Now question is arise how the news impact can be used in stock Price prediction, basically automatic text classification techniques are used to analyze the incoming news.

The aim of this paper is to develope prediction system to improve the accuracy of stock market price prediction by using combine approach of rule based classification, technical analysis and effect of financial news also. The rest of the paper consist following sections as followed. Section 2 presents a review of related work.Section 3shows the impact of financial news in stock price prediction. Section 4 introduces the basic concept of classification rules and its uses for prediction the stock price. Section 5 describes our proposed work and methodology we used while Section 6 define result. And last section conclude the paper.

II.   REVIEW OF LITERATURE

Jageshwer Shriwas et.al [2] develop a model that would accurately predict the future closing price of the BSE. Once this was accomplished, the probability of an accurate forecast would be calculated. Given the accuracy of the forecast, the benefits of the network to the investor would be determined. Financial time series consists of multidimensional and complex nonlinear data that result in of pitfalls and difficulties for accurate prediction. This system use neural network as a prediction tool and accuracy is compared against a traditionally forecasting method.

Anil Rajput et.al.[12] focuses for the rule based classification model of historical BSE stock data with data mining techniques. They used decision tree and rule induction method with the help of data mining software. They have to create classification rules and induction rules with the use of J48 and PRISM classifiers under WEKA software. Stephen Evans[6] uses classification data mining to attempt to predict the direction of daily returns of randomly selected stocks from the Russell 1000 and Russell 2000 stock indexes.The study uses moving averages of historical daily stock prices as attributes, along with different data mining classifiers, to attempt to make these predictions. A secondary goal of this study is to determine how effective using Distributed Data Mining (DDM) can be in predicting the direction of daily stock returns. Hence, DDM classifiers are used in the testing. This study discovers that the moving averages of daily returns do not help predict the direction of future daily stock returns any better than the percentages of returns from one trading day to the next. It also shows that the classifiers were no more than 60% accurate in predicting the directions of daily returns for any of the stocks used in this study.

S. Karthik et.al.[16] objective is to analyze the stock market trend using integrated clustering and weighted rule mining technique. For predicting and analyzing the market trends data mining techniques are also used. Statistical techniques are used for the market price prediction process. Inaccurate results are produced in the statistical analysis. To find the stock market trend with index dependency analysis environment hybrid clustering and association algorithm is not appropriate. Statistic analysis techniques are not suitable for trend analysis with index relationship. The stock market transactions data is analyzed with clustering and weighted rule mining techniques. The K-mean clustering algorithm is used to cluster the transaction with respect to the market flows. The market trade transactions are divided into three zones such as up trend, down trend and stable zone. The weighted rule mining technique is applied to fetch patterns from the indexes, sector indexes and company price values. Apriority algorithm is modified to carry out weighted rule mining process. The system produces the market trade trend flow with market indexes and sector index values.

S. Prasanna et al.[19] have discussed several attempts made by researches for stock price prediction. These works show that data mining techniques can be applied for evaluation of past stock prices and acquire valuable information by estimating suitable financial indicators. They discussed works basically two types of prediction methods are implemented by several researches to generate useful extracts. They are fundamental approaches and technical indicator based approaches. Many researchers adopted technical indicator approaches only. Limited work is done with fundamental approaches which give plenty of opportunity for further research. Schumaker, R. P.[11] examines a predictive machine learning approach for financial news articles analysis using several different textual representations: bag of words, noun phrases, and named entities. The proposed model called AzfinTS based on

542

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

support vector machine approach. It calculate the stock price after 20 minute to released news. The model containing both article terms and stock price at the time of article release had the best performance in closeness to the actual future stock price, the same direction of price movement as the future price and the highest return using a simulated trading engine. Proper Noun scheme performs better than the de facto standard of Bag of Words in all three metrics.

Manisha V. Pinto et.al.[20] provide a framework for predicting stock magnitude and trend for making trading decisions by making use of a combination of Data Mining and Text Mining methods. The prediction model predicts the stock market closing price for a given trading day _D', by analysing the information rich unstructured news articles along with the historical stock quotes. In particular, we investigate the immediate impact of the news articles on the time series based on Efficient Market Hypothesis (EMH).Key phrases provide semantic metadata that summarize and characterize documents. This framework incorporates Kea [22], an algorithm for automatically extracting key phrases from news articles. The prediction power of the Neural Network is used for predicting the closing price for a given trading day. The Neural Network is trained on the extracted key phrases and the stock quotes using the Back propagation Algorithm. Now a days, traders need to use various prediction techniques parallel to achieve more information about the future of the markets.

III.   FINANCIAL NEWS THEIR APPLICATION IN STOCK PRICE PREDICTION

Financial News based stock market prediction can be considered as a textual classification task. Generally the aim   is to forecast some aspects of the stock market such as price or volatility based on the news content. Information in the form of quotes and financial news is released into the market all the time. While quotes data is structured and can be directly used, the challenge is to manage the large amounts of textual information. We can employ techniques to parse the news articles and identify the key features most likely to have an impact on the stock market. By automating this process, machines can take advantage of arbitrage opportunities faster than human counterparts by repeatedly forecasting price fluctuations and executing immediate trades to make profits[20].At a time lots of news articals are released.We know that news articles are an unstructured form of data with enormous amount of useful information embedded in it. KEA [22], an algorithm for automatically extracting key phrases from text is used[20,22].The most common key phrases extract from the news articals which influence the stock prices are listed below:

TABLE 1: SHOWS TOP 20 KEY PHRASES [20]

'opens lower'
'jobless claims'
'open higher'
'slumping dollar'
'battered financial'
'oil futures closing high'
'unemployment rate'
'oil prices'
'lifted financial shares'
'credit crisis'
'strong gains'
'jobless claims data'
'drop in crude oil futures'
'economic news'
'investors\' confidence'
'quarterly loss'
'nuclear power'
'investors\' appetite'
'nuclear crisis'
'required reserves'

543

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Based on prediction goal a set of final classes are defined, such as ―Up‖(which means this news cause the prices to go up), ―Down‖ (which means this piece of news is probable to causes decrease in prices) and etc. the prediction system is supposed to classify the incoming news into one of these classes. News based market prediction can be divided into two main phases. ―Training phase‖ and ―Operational phase‖. In operational phase, one of the predefined classes will be assigned to incoming news; however, to make the system ready for the operational phase a classifier should be trained in the training phase. Machine learning techniques are widely used to automate such processes. As a part of the training phase, a set of training data shall be prepared which in our case the train data are the pre-classified news and market information such as market prices. These labeled (pre-classified) news and possibly market numerical data will be processed to be fed into the classifier for training. The trained classifier would be ready to get a piece of news and assign a class to it in operational phase. The accuracy of the system is measured as the percentage of the predictions that were correctly determined by the system. For instance, if the system predicts an uptrend and the index indeed goes up, it is assumed to be correct, otherwise, if the index goes down or remains steady for an uptrend, it is assumed to be wrong.[20,22].Representing the news section consists of two main tasks:feature selection and feature weighting. First a set of features will be selected to represent a piece of news and next step is to assign weights to theses selected features. These weighted vectors would be the inputs to classifier. In real-world trading applications, the amount of textual data available to stock market traders is staggering. This data can come in the form of required shareholder reports, government-mandated forms, or news articles concerning a company's outlook. Reports of an unexpected nature can lead to wildly significant changes in the price of a security.

IV.   CLASSIFICATION RULES AND THEIR ROLE IN STOCK PREDICTION

Classification is one of the approaches of Data Mining Techniques.Classification and prediction is the process of identifying a set of common features and models that describe and distinguish data classes or concepts. The models are used to predict the class of objects whose class label is unknown. A bank, for example, may classify a loan application as either a fraud or a potential business using models based on characteristics of the applicant. A large number of classification models have been developed for predicting future trends of stock market indices and foreign exchange rates[4].The objective of classification is to first analyze the training data and develop an accurate description or a model for each class using the features available in the data. Such class descriptions are hen used to classify future test data in the database or to develop a better description (called classification rules)for each class in the database[17]. Classification is perhaps most familiar and most popular data mining technique. classification application include image and patter reorganization, medical diagnosis, loan approval detecting faults in industry application and classifying financial market trends. Estimation and prediction may be viewed as prediction a continuous values, while classification forecast a discrete values. Classification are categories in many groups. These are Statistical based, Distance based, Decision tree, NN and Rule based. Rule based classification are used to define if-then rules that can be used to explain the result of given training sets[5,6].

V.   PROPOSED WORK

Some prediction methods used historical stock prices and use technical analysis to predict the stock market. Some other methods are based on rule based classification. In this paper, the method we used based on combine approach of ruled based classification and technical analysis to developed stock prediction system .

A.   Method

Pattern matching using Rule Based Classification and Technical analysis to predict thefuture movement direction of a index values. After determining the future movement direction, severa lformulas are applied to calculate the index value for the next trade date. The logical flow of the whole financial forecast process is shown in Figure 1.
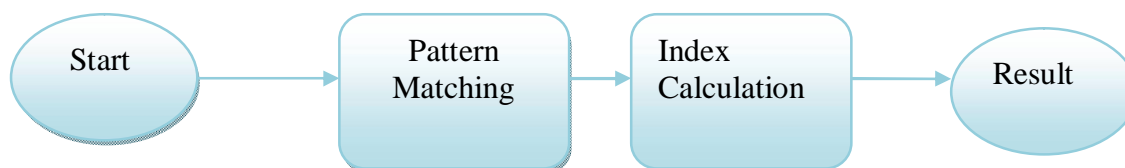
Start → Pattern Matching → Index Calculation → Result

Figure 1: Logical flow of financial forecast process.

B.   Rule Based Classification

Rules are a good way of representing information or bits of knowledge.A rule based classifier uses a set of if-then rules for

544

classification.Rule-based classification is a type of supervised learning data mining method[11]. A key approach to generate subsets from given data set through if-then rules that cover all the appropriate conditions. Expression of if-then rule is:-

If condition then conclusion

For example:-

R1 : If age=youth AND student=yes then buys-comp=yes

The if part (or left side) of a rule is known as the rule antecedent or precondition.The then part (or right side) is the rule coBSEquent.In the rule antecedent, the condition consists of one or more attribute tests (eg age and student) that are logically ANDED.The rules cosequent contains a class prediction (in their case we are predicting whether customer will buy computer buys-comp is yes).R1 also written as:

R1: (age=youth) ^(student=yes) => (buys-com=yes)

If the condition(i.e. all the attribute tests)in a rule antecedent holds true for a given tuple,we say that the rule antecedent is satisfied(or simply that the rule is satisfied) and that the rule covers the tuple.A rule R can be assessed by its coverage and accuracy.Given tuple ,X,from a class-labeled data set,D,let ncovers be the number of tuples coverd by R; nconnect be the number of tuple correctly classified by R; and $|D|$ bethe number of tuples in D.We can define :

Coverage (R)= ncovers / $|D|$

Accuracy (R)=ncorrect / ncovers

That is,a rules converage is the percentage of tuples that are covered by the rule(i.e.their attribute values hold true for the rules antecedent).For a rules accuracy we look at he tuples tha it covers and see what percentage of them the rule can correctly classify[1]. In our paper this rule is used to find out the patter of coming financial news from different source into predefined class Up (which means this news cause the prices to go up) and Do (which means this piece of news is probable to causes decrease in prices).

C. Technical Analysis

Technical analysis is the method of predicting the appropriate time to buy or sell a stock. The idea behind Technical analysis is that the share prices move in trends dedicated by the constantly changing attributes of investor. Using technical data such as price,volume,highest and lowest prices per trading period. Technical analyst believes that share prices are determined by the demand and supply forces operating in the market. These demand and supply forces in turn are influenced by a number of fundamental forces as well as certain psychological or emotional factors. Many of these factors cannot be quantified. The combined impact of all these factors is reflected in the share price movement. A technical analyst therefore concentrates on the movement of share prices. He claims that by examining past share price movements future share prices can be accurately predicted. The basic premise of technical analysis is that prices move in trends or waves which may be upward or downward. It is believed that the present trends are influenced by the past trends and that the projection of future trends is possible by an analysis of past price trends. A technical analyst, therefore, the price and volume movements of individual securities as well as market index. Thus, technical analysis is really a study of past or historical price and volume movements so as to predict the future stock price behavior[17].The technical analyst uses charts to predict future stock movement .Price charts are used to detect the trends or daily movement of market. These trends and movements are assumed to be based on supply and demand issues which often have cyclical or noticeable patterns.

D. Index Calculation

A number of formulas are used to calculate the predicted index value for the next trade date after determined the movement direction, and their performances are being assessed by comparing to the differences to the actual index point.The predicted index point is generally calculated by the last index point plus a movement. The movement direction predicted determines the impact of financial news in stock price. Formula is:

$$Y_{n+1} = Y_n \pm (\sum V_j) / n,\ 1 <= i <= n$$

Where $Vi = y_n - y_n - 1$ // movement magnitude

E. Proposed Algorithm
STEPS FOR ALGORITHM IS AS FOLLOWS

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

STEP 1:   Collect all the index values as a data set.

STEP 2:   Data set comparison

                 Compare first index and next index value.

STEP 3:   Store it into movement

STEP 4:   Find all movement values

                 With compare all data set store it to movement sum.

STEP 5:   Checking for up or down

                 If average movement sum is positive then   //positive denotes next index value is more than previous

                 Count Future index as a up.

                 If average movement sum is negative then   //negative denotes next index value is less than previous

                 Count future index as a down.

STEP 6:   Count the percentage future index values.

STEP 7:   Exit

VI.   RESULT

Sample index values were collected from internet.For 15 prediction we compared the performance of each index calculation. The system predicted the Predicted index movement direction successfully for 12 out of 15 samples, the accuracy was 81.0%. Graphical representation of actual index value and predicted index value of our system.
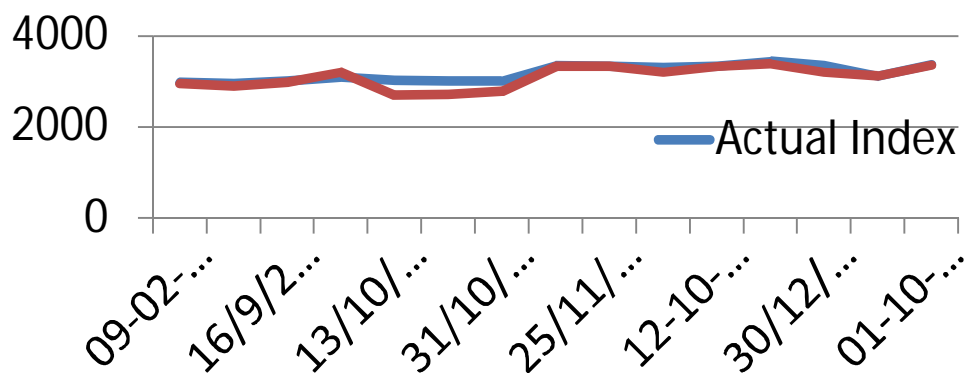

Figure 2: Graphical representation of actual and predicted index values on daily basis.

VII.   CONCLUSION

This paper presents the predicted system used data of BSE to predict the stock market trend. This system used hybrid approach of technical analysis and rule based classification of data mining. Simultaneously we check the effect of news article to daily stock price also.First we find out the pattern of incoming financial news and then calculate index value from formula. News input could increase performance in most cases or at least maintain the performance of the current system. In order to achieve better results, you would rather want to combine classical trading strategy with one based upon real-time technical indicators.

VIII.   ACKNOWLEDGMENT

REFERENCES

[1]   Jageshwer Shriwas, Shagufta Farzana," Using Text Mining andRuled Based TechniqueFor Prediction of Stock Market Price" International Journal of

546

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Emerging Technology and Advanced Engineering (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 1, January 2014

[2]   Jageshwer Shriwas, Dr Samidha Dwivedi Sharma," Stock Price Prediction Using Hybrid Approach of Rule Based Algorithm and Financial News" Int.J.Computer Technology & Applications,Vol 5 (1),205-211

[3]   Margaret H.Dunham,‖Data Mining Introductory and Advanced Topics.‖ Impression,2012.

[4]   Jiawei Han/M.Kamber/Jjian Pei,Data Mining Concepts and Techniques.Impressio 2012.

[5]   Dongsong Zhang and Lina Zhou,"Discovering Golden Nuggets: Data Mining in Financial Application",2004

[6]   Stephen Evans,"Data Mining in Financial Markets",.2011.

[7]   Sikora, Riyaz and Michael Shaw. "A Computational Study of Distributed Rule Learning."Information Systems Research 7.2 (1996): 189-197.

[8]   TanviChauhan, Prof.VineetRichhariya, Sunil Sharma, "Literature Report on Face Detection with Skin &Reorganization using Genetic Algorithm", TanviChauhan et al. / IJAIR Vol. 2 Issue 2 ISSN: 2278-7844 .

[9]   Lo, A.W., & MacKinlay, A.C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test, Review of Financial Studies 1, 41-66.

[10]   Deng, Min. Pattern and Forecast of Movement in Stock Price. [Online] Available: http://ssrn.com/abstract =217048 or DOI: 10.2139/ssrn.217048 (November 8, 2006)

[11]   Schumaker, R. P. and Chen,Textual analysis of stock market prediction using breaking financial news: The AZFin text system. ACM Trans. Inform. Syst. 27, 2, Article 12,February 2009.

[12]   Anil Rajput , S.P. Saxena , Ramesh Prasad Aharwal and Ritu Soni ―Rule based Classification of BSE Stock Data With Data Mining‖. International Journal of Information Sciences and Application.ISSN 0974-2255 Volume 4, Number 1 (2012).

[13]   R Miri, P Tripathi, S R Tandan "Novel Algorithm for Reduced Computational Data by Using Fuzzy Classification and Data Mining Techniques" Proceedings of the 2014 International Conference on Information and Communication Technology for Competitive Strategies" ACM, (2014)

[14]   Mary K. Obenshain, MAT (2004): Application of Data Mining Techniques to Healthcare Data, Statistics for Hospital Epidemiology,(25),(8), 690-695.

[15]   Samuel Odei Danso‖ An Exploration of Classification Prediction Techniques in Data Mining: The insurance domain‖, September, 2006 .

[16]   S.Karthik , K.K.Sureshkumar ," Analysis of Stock Market Trend using Integrated Clustering and Weighted Rule Mining Technique", International Journal of Computer Science and Management Research, ISSN 2278-733X, Vol 1 Issue 5 December 2012.

[17]   Ming-Syan chen,Jiawei Han,Philips S.Yu,,"Data Mining : An overview from a Database Perspective", IEEE Transactions on Knowledge and Data Engineering,Vol 8,No 6 December 1996.

[18]   S.Kevin,"Security Analysis and portfolio Management" ,PHI learning, Edition 1,2006.

[19]   S.Prasanna , Dr.D.Ezhilmaran,‖An analysis on Stock Market Prediction using Data Mining Techniques‖, International Journal of Computer Science & Engineering Technology (IJCSET).

[20]   Manisha V. Pinto, Kavita Asnani" Stock Price Prediction Using Quotes and Financial News" International Journal of Computer Science & Engineering Technology (IJCSET)2011.

[21]   Azadeh Nikfarjam, Ehsan Emadzadeh ,Saravanan Muthaiyah," Text mining approaches for stock market prediction",IEEE 2010.

[22]   Ian H. Witten,Gordon W. Paynter,Eibe Frank,Carl Gutwin and Craig G. Nevill –Manning, "KEA: Practical Automatic Keyphrase Extraction" .

[23]   http://www.learnartificialneuralnetworks.com/stockmarketpr ediction.html.

547

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ◎ (24*7 Support on Whatsapp)