An Efficient Prediction of Share Price using Data Mining Techniques

Priyanka Garg, Santosh K. Vishwakarma

Abstract-The prediction of share prices is the function of deciding the future price of a company stock or other commercial tool traded. Prediction of some movements allowed from some patterns can be found. People are always attracted to invest in share market and stock exchanges as they provide huge financial profits, which is also an important for finance research. Prediction of share price is very difficult issue it depends upon such huge numbers of factors such organization financial status and national policy and so on. Nowadays stock costs are influenced because of numerous reasons such as organization-related news, political, socially efficient conditions and cataclysmic events. Many studies have been performed for the prediction of stock index value and daily direction of change in the stock index. Such huge numbers of models have been created for foreseeing the future stock costs yet everyone has their own weaknesses. This paper expects to study, develop and assess different techniques so as to foresee future stock trades. The experimental results states that different classification techniques can be successfully deploy for share price prediction.

Index Terms—Classification Algorithms, Deep Learning, Naïve Bayes, Rapidminer, Share Price.

I. INTRODUCTION

Predicting return on the stock market is a significant problem in financial institutions and also it isa complicated issue. The stock price prediction was always a difficult task. It has been seen that any company's stock costs don't really depend solely on the money related status of the company, yet additionally on the country's social economic situation. It is no longer directly related to the country's particular area's economic development. So today's shareprice prediction has turned out to be much harder than previously. Because of many reasons such as political events,company related news, natural disasters share prices are affected these days.

A lot of research in forecasting stock prices or stock index has been going on for many years. It involves assuming basic information that has been freely accessible in the past that has certain anticipatingrelationships with future stock returns or records. The share value depends on how many people are interested in buying it and how many people are selling it.

The price will rise if a lot of people want to buy a stock. The cost will drop if there are a bigger number of sellers than buyers. Usually people buy/sell stock shares with a broker's help. A broker also assists client in making good stock indices. For most of the stocks, most brokers have recommendations based on company information and what is expected them.

The efficient market theory indicates that share prices mirror all currently available information and therefore any price changes not based on newly disclosed information are inherently unpredictable.Stock markets are an ever changing chaotic business area where prediction plays an important role.

Prediction provides information on the current status of the movement of stock prices. This can therefore be used in customer decision making to finalize whether to purchase or sell the specific shares of given stock.

To formulate future predictions, predictive analysis uses historical data. These predictions rarely acquire the structure of absolute results, and are described to show the behavior corresponding to future behavior.

Classification is a techniqueof data miningused to allocate each item in a set of data into one of a pre define set of classes of group the classification technique is used. The mathematical techniques such as decision tree, linear programming, neural network and statistics are used by classification method.

II. REVIEW OF LITERATURE

Research by Zhao, Lei and Wang, Lin [1]figured out how to find the irregularity trading data between the high-recurrence tick-by-tick information by looking at the volume sequence circulation between market and stock. They found that anomaly clusters dependably foresee an upward pattern in the stock cost.The authors of [1] tested outlier data mining algorithm and found that it is reliable with clustering algorithm (k-means). The normal return and effective rate is examined against outlier algorithm and the forecast about these two amounts is right and satisfactory.

The authors SurajBaviskar and NitinNamdev[2] centered on understanding inventory marketplace related facts for buyers and analyzed National Stock Exchange information to are expecting destiny inventory actions. Data mining has been drastically used to extract important records from historical stock data to research and are expecting future traits.

Retrieval Number F9085088619/2019©BEIESP
DOI: 10.35940/ijeat.F9085.088619

3110

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

There are two categories of technical indicators used to better understand stocks: First is a price-based indicator and second is a volume-based indicator. And these two indicators define different algorithms.

Weerachart and N.Benamas[3] presented a predictive model that uses Data Mining techniques to forecast share price trends.The author used the Gain Ratio Attribute in this study to compare the efficacy of feature selection with the Ranker Search Method and Wrapper Selection using Greedy Step Wise Search Method.By using Wrapper Subset Evaluation with Greedy algorithm through forward selection, the attributes are decreased from 14 to 6 (57.14 percent).The outcomes of this study show that the predictive model for the direction of weekly stock price is improved by using ANN classification, where the maximum accuracy of the model reached 93.89 percent, which was a major improvement in the daily and 5-day predictions using only six chosen attributes.

Qasem,Assaf and Alnagi [4] used the decision tree classification, and it's one of the techniques of data mining.The CRISP-DM methodology is used to construct the proposed model. This study presents a proposal to use the decision tree classifier on stock historical prices to create decision rules that give stock market buy or sell recommendations. The results for the proposed model were not perfect because many factors influenced the stock market, including but not limited to political events, general economic conditions, and expectations of investors.

Using historical time series stock market data and data mining techniques, the authors Mahntesh C. Angdi and Amogh P. Kulkarni[5] developed a prediction model for predicting share market patterns depending on technical analysis.The experimental results obtained demonstrated the potential of ARIMA model to predict the stock price indices on short term basis. A stock market shows investments and savings that are beneficial to enhance the national economy's effectiveness.R is a language of programming and a graphics and processing environment. R-studio allows the user to run R scripts in a more user friendly environment.

The authors G.S.Navale,Nishant,Kunal,Pawan and Brijkishor[6] used artificial intelligence and data mining to accurately predict the results. In artificial intelligence, most researchers have used methodologies to achieve accuracy and results. And the parameters and performance were improved. Using data mining, this can be achieved.For accurate result they used data mining and artificial intelligence together. The authors [6] have used Auto regressive moving average algorithm for prediction.

Ruchi Desai and S. Gandhi [7] utilized Text mining way to deal with measure the impact of continuous news on stock. They exhibited a model that predicts changes in stock pattern by examining the impact of non-quantifiable data, specifically news stories that are affluent in data and better than numerical information. This technology is developed to help shareholders discover hidden patterns from historical data that are likely to forecast their investments.

S Prasanna and D. Ezhilmaran [8] searched for work on the prediction of stock prices. In this review, the author tried to narrate some important work done using data mining techniques to predict stock prices. This survey is shortly explaining the works done for stock price prediction.

The writers (Dr. P.KSahoo and Mr. Krishna Charlapally) of[9] explored the use of auto-regressive approach to forecast stock rates. Due to its simplicity and broad acceptability, the design of auto regression is used. They also performed a research on the auto-regressive model's efficacy. The method of Moore and Penrose is used to predict the regression coefficient coefficients. And they also studied prediction accuracy by comparing the predicted values over a period of time with the actual values.

Mr. Pritam R Charkha [10] tested Feed Forward Network using Early Stopping Back Propagation Learning and Radial Basis Neural Network to forecast inventory market trends and inventory cost predictions.In this study, fundamental information or technical indices were not used as the fundamental goal of this study was to determine the usability of artificial neural networks in anticipating potential rates based solely on previous rates.

III. METHODOLOGY

The purpose of our paper is to use different data mining technique to predict future share price of State bank of India. We will use different data mining algorithms to predict share price and proposed different ways to combine the results of different algorithms. This research has been done by collecting historical share price of SBI from yahoo finance website.

Below flowchart shows

Stock trading data
↓
Preprocessing
↓
Apply data mining techinque
↓
Result
↓
Prediction
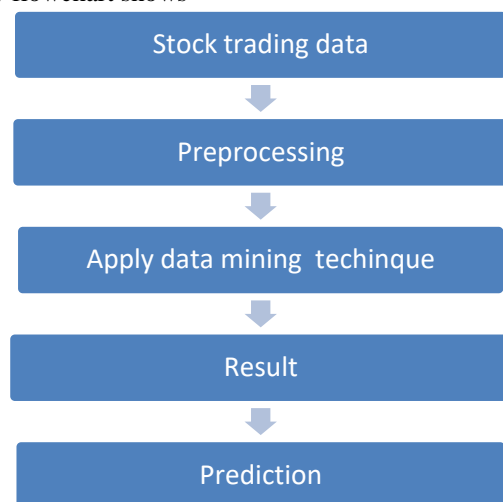
the step by step final process:
Fig. 1.Steps involved in prediction

This system includes the following steps-

1. Data collection- Data collected from yahoo finance website from 16march2018 to 24 October 2018 in excel.
2. Data preprocessing- Modify downloaded and preprocessed.
3. Apply data mining technique- Apply classification technique. For better result we apply different data mining algorithms i.e. deep learning, generalized linear model, Random Forest, Decision tree, K-nn and Naïve bayes.

4. Result- shows different attribute such as accuracy, kappa, classification error, future stock value predicted by all model.
5. Prediction- Combine the result of these applied model, compare them based on its accuracy and predict best model result.

SBI share trading data- It is very essential to know about the share trading data of company which is used for prediction process. This past data is easily accesses from website. Snapshot show the downloaded past share prices of SBI.

Fig. 2.SBI historical share price.

Proposed system Model:-
This system will predict and also define different attributes of process. A new approach for forecasting future share price is proposed by data mining classification. This idea is extracted from the technical analysis. 150 days data transformed into a new data set, which consist of 7 attributes (5predictor attributes and 2 target attributes).

Fig. 3.Transformed data set.

We prefer to use Rapid miner 8.1 tool which is open source software. By this we can easily apply different classification algorithm on our past stock data. We apply one by one all operators in our process. These operators are Deep learning, Naïve bayes, GLM and K-NN. Below figure 3 shows the all four snapshot of our final process in RapidMiner.
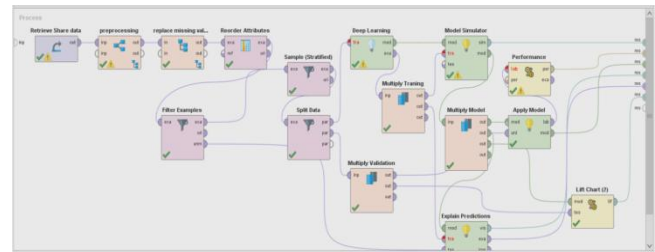

Fig.4 (a).Applying Deep Learning operator
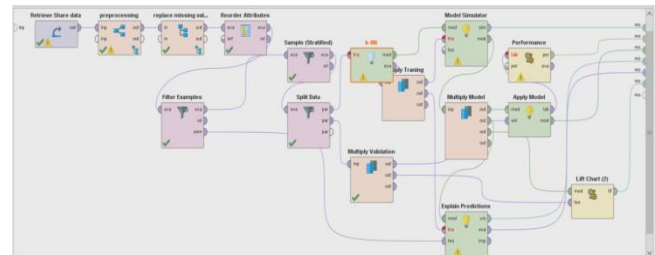
Fig. 4(b).Applying K-NN operator.
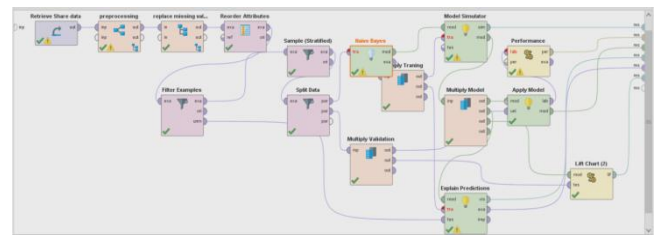
Fig.4(c).Applying Naïve Bayes operator.
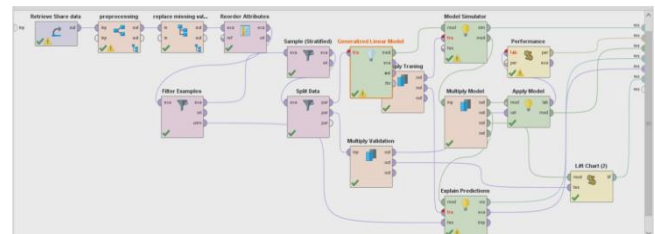
Fig. 4(d).Applying GLM operator

In this process STATUS attribute set as a label attribute. Here we apply one by one all four operators. In RapidMiner it is very simple to replace the operator. Figure 4 showing the result of process. Here we can see that the model gives prediction of next day price and shows status positive. In binary classification, classification algorithm classifies a model in two ways that is positive or negative.

Fig.5. Result of the process.

Below snapshots shows the Performance vector of all four model.The Performance Vector is a list of efficiency criteria values. Calculation of the performance vector is focused on the label attribute and prediction attribute of the Example Set. The performance operator will calculate the output-performance-vector. The output-performance-vector includes performance criteria that the performance operator calculates.To evaluate the model, various performance measures are used:

accuracy: 90.00%

	true neg	true pos	class precision
pred. neg	14	1	93.33%
pred. pos	2	13	86.67%
class recall	87.50%	92.86%	

Fig. 6.Performance of Deep Learning operator.

accuracy: 60.00%

	true neg	true pos	class precision
pred. neg	6	2	75.00%
pred. pos	10	12	54.55%
class recall	37.50%	85.71%	

Fig. 7.Accuracy of Naïve Bayes operator.

accuracy: 86.67%

	true neg	true pos	class precision
pred. neg	14	2	87.50%
pred. pos	2	12	85.71%
class recall	87.50%	85.71%	

Fig. 8.Accuracy of GLM operator.

accuracy: 80.00%

	true neg	true pos	class precision
pred. neg	12	2	85.71%
pred. pos	4	12	75.00%
class recall	75.00%	85.71%	

Fig. 9.Accuracy of K-NN operator.

These figures (fig. 6, 7, 8, 9) give the various criteria in result workplace. These shows accuracy, classification error, class recall and class precision.

IV. RESULT & ANALYSIS

Prediction of SBI is done by using 151 days historical share price which is easily downloaded from Yahoo finance website. In this paper, the open source data mining tool RapidMiner is used for implementation. These predictions are automatically made by the Rapidminer tool. Different classification models such as Naive Bayes, K-NN, Deep Learning and Generalized linear model are applied after pre-processing in the dataset and a comparative study was carried out based on their performance and predictive accuracy.

This table (Table I) gives the result about various attribute selected for stock data analysis. Result shows that Deep learning is the best model for stock data analysis in comparisons with other models. Accuracy of deep learning is 90% that is higher than other models. The attributes that shows in table also define below:

Table-I Performance table			
Model	Accuracy	Classification error	Kappa
Naïve bayes	60%	40%	0.224
GLM	86.67%	13.33%	0.732
K-NN	80%	20%	0.602
Deep Learning	90%	10%	0.800

Accuracy-Accuracy is a total number of examples properly classified, and in other words it gives predictions percentage correctly.

Classification_error-It is a total number of wrongly classified examples and, in other words we can say it is a percentage of wrong predictions.

Kappa-Kappa classification data are usually believed to be a more solid measure than a easy calculation of a percentage-correct forecast because it takes into consideration the accurate forecast that occurs by chance.

The following table shows the comparative study for different classification algorithms together with their performance parameters such as accuracy, classification error rate and Kappa.

Under the converted data set, the prediction model's maximum accuracy was 90% using only seven attributes over a period of 150 days. From the above table, we can see that Deep Learning Model has delivered the best performance in terms of achieving high accuracy and low classification error rates.
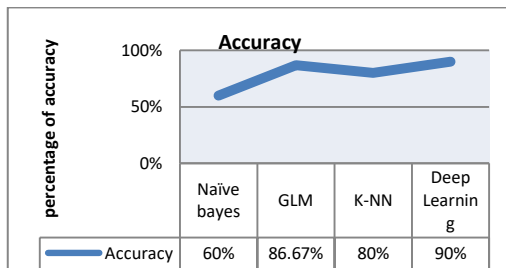
Fig. 10.Graph of accuracy

Figure 10 shows the graph of accuracy of all models which is used for prediction. Accuracy is defined as how the system gives the correct prediction. Above accuracy graph shows that Deep Learning model achieved highest accuracy of 90%. Here we can also see that other classifiers accuracy rate where Naïve bayes classifier gives lowest accuracy rate.
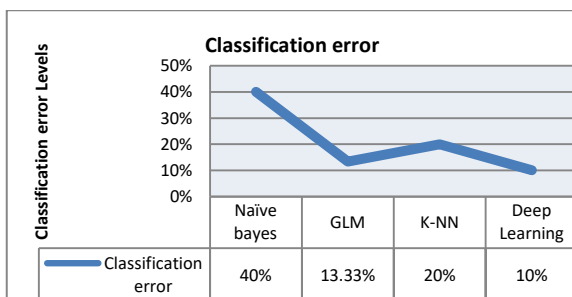

Fig. 11.graph of classification error

The classification error percentage is the percentage of wrong predictions.The above graph indicates that Deep Learning Model has given the least 10 percentage classification error rate as it has achieved a high accuracy rate. If we compare to other classifiers we see that Naïve bayes classifier given high error rate as it gained low accuracy.
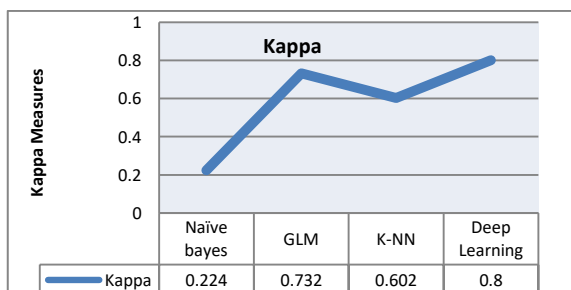

Fig. 12.graph of Kappa classification

Kappa statistics is a prevalent measure of how the system makes the predictions correct.Compared to accuracy, it is considered more reliable; as it also considers correct predictions made by chance by our Rapidminer system.It compares process prediction accuracy with a random process.The analysis indicates that the highest Kappa value of Deep Learning Model is 0.800 which is close to 1. Other classifiers also gained high values of Kappa.

V. CONCLUSION

In order to analyze and predict future trends, data mining classification has been usedwidely to extract essential information from historical stock data.This study presents a proposal to use the various data mining classifier on stock historical prices which give recommendations for buying or selling in the stock market. Such proposed system can be a useful tool for shareholders to make the correct choice about their shares based on the analysis of stock historical prices to extract any predictive information from that historical data. By using the Rapidminer tool, we examined this research and applied different prediction models techniques. We compared different predictive functions and found that Deep Learning function provides the ability to predict SBI stock price more accurately than other functions such as Naïve bayes, Generalized Linear Model, K-NN. This analysis can be used in predicting future share prices to reduce the percentage of errors. It increases investors ' chances of predicting prices more accurately by reducing the percentage of error and thus increasing share market profit. WE can find more accurate results if we take longer duration data. It will help to check value variations and predict more accurate outcomes.

Future work can be done, by applying more techniques and verifying results with actual values. We can design the error rate and graph to study the validity of it. After this step, it is possible to select the best technique by observing error rate on different methods. With different data set, all of these steps can be repeated.

REFERENCES

1. Zhao, Lei Baylor University,Wang, LinJapan Advanced Institute of Science and Technology "Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm" in2015 IEEE Fifth International Conference on Big Data and Cloud Computing CFP1552Z-CDR/15 ©2015 IEEE,DOI 10.1109/BDCloud.2015.19.
2. S.Baviskar,N.Namdev"Analyzing and Predicting Stock Market Using Data Mining Techniques" August 2015,IJIRT Volume 2 Issue 3,ISSN: 2349-6002.
3. W.Lertyingyod and N. Benjamas "Stock Price Trend Prediction using Artificial Neural Network Techniques" in IEEE 978-1-5090-4420-7/16/2016 IEEE.
4. Q.A.Al-Radaideh, A. A. Assaf, E. Alnagi, "Predicting Stock Prices Using Data Mining Techniques" inThe International Arab Conference on Information Technology (ACIT'2013).
5. M. C. Angdi, A. P. Kulkarni, "Time Series Data Analysis for Stock Market Prediction using Data Mining Technique with R", in International Journal of Advance Research in Computer Science, Volume-6, No.6, July-august 2015,ISSN No.0976-5697.
6. G.S.Navale, NishantDudhwala, KunalJadhav, PawanGabda, BrijKishorVihangam, "Prediction of Stock Market using Data mining and Artificial Intelligence", in International Journal of Computer Applications(0975-8887)Volume134-No.12,jan2016 SavitribaiPhule Pune University, SITS Nahre, Pune-411041.
7. R.Desai,Prof. S. Gandhi, "Stock Market Prediction Using Data Mining", 2014 IJEDR(International Journal of Engineering Development and Research) Volume 2, Issue 2 ISSN: 2321-9939.
8. S. Prasanna,Dr. D. Ezhilmaran, "A Survey of Stock Price Prediction & Estimation Using Data Mining Techniques", in International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 6 (2016) pp 4097-4099.
9. Dr. P. K. Sahoo, Mr. Krishna charlapally, "Stock Price Prediction Using Regression Analysis" , in International Journal of Scientific & Engineering Research, Volume 6, Issue 3, March-2015 ISSN 2229-5518.
10. Mr. P. R. Charkha, "Stock Price Prediction and Trend Prediction using Neural Networks" in 2008 IEEE DOI 10.1109/ICETET.2008.22.

AUTHORS PROFILE

PriyankaGarg graduated in Information Technology from Gyan Ganga Institute of Technology and Sciences Jabalpur,M.P.,India. At present, she is pursing Master of Technology in Computer Science and Engineeringfrom Gyan Ganga Institute of Technology and Sciences Jabalpur.Her area of interest includes data mining, Machine Learning and Text Mining.

Santosh K. Vishwakarma has completed his PhD, Masters and Bachelor's degree in Computer Science & Engineering. Currently employed with G.G.I.T.S. Jabalpur as Professor, CSE and Dean – Research & Entrepreneurship Cell. He has organized various International Conferences, Workshops, Seminars and Entrepreneurship Awareness programs. His research interest includes Machine Learning, Data Mining and Text Mining. He has several International and National publications in his profile. Currently he is actively involved in IEEE MP Sub section as Execom member & Student Activity Chair. He is also appointed as the Secretary, Computer Society of India – Jabalpur Chapter.

3115