Stock Market Trend Analysis

Dr. Devpriya Soni[1], Sparsh Agarwal, Tushar Agarwal,
Pooshan Arora[2], Kopal Gupta
Department of CSE, Jaypee Institute of Information Technology,Noida , Uttar Pradesh, India
E-mail: [1] devpriya.soni@jiit.ac.in ; [2] pooshanarora19@gmail.com

***Abstract*— The main objective of this work is to add to the academic understanding of stock market analysis using some well defined algorithms and machine learning techniques. Stock price forecasting is a popular and important topic in financial studies and at academic levels. Share Market is not a neat place for analyzing since there are no significant rules to estimate or predict the price of share in the share market. Many a method like technical analysis, fundamental analysis, time series analysis and statistical analysis, etc. have been used in an attempt to analyze the share trends in the market but none of these methods have so far proved to be a universal approach for acceptance as a prediction tool.**

The intricacy while analyzing market trends is that they have a dependency on a number of external factors ,some of which are not under one's control. The goal of this work is to analyze stock market trends using some machine learning techniques. After analyzing the trends with the help of standard techniques , we then proposed an entirely new approach to analyze stock market indices over which accuracy is calculated and compared over different techniques and algorithms. We outline the design of the proposed model with its salient features and customizable parameters. We finally tested our model on the one year of Nifty stock index dataset at real time where we analyzed the values on the basis of data from the past days for three months.

I. INTRODUCTION

Predicting the Stock Market has been tiring and troublesome for investors since market's existence. Share Market is an area where prediction doesn't follow any significant rules to analyze the price of a share in the share market. Both investors and industry are involved in stock market and want to know whether a particular share would rise or fall over a certain period of time. The share market is based on the concept of demand and supply. If the demand for a particular company's stock is higher, then that company's share price would tend to increase and if the demand for company's share is low then the company share value tends to decrease.

Another motivation for research in this field is that it possesses and poses many theoretical and experimental challenges to researchers, some of which cannot be determined directly. Companies rise and fall daily based on the behavior and existing trends of the market .One such interesting and significant challenge is the Efficient Market Hypothesis (EMH). Since there are a large number of people competing to analyze market trends and predicting a share's value and since almost every kind of relevant information needed for the same is readily available, an individual is led to a situation where share market values readily reflect available information about the existing market trends and thus an opportunity to gain extra profit does not exist. Due to the presence and involvement of a number of companies and industries, extremely large data sets prevail from which it is difficult to decipher information and analyze their working trends manually or without external help. Share market analysis and prediction depicts and reflects the market patterns and gives an idea of the time period to purchase a particular stock. The successful prediction of a stock's price by its analysis could lead to a significant profit. This is being done with the help of extremely large historic data sets to depict varying conditions and thus reaffirming the belief that the time series patterns possess significant predictive power with a high probability to generate profitable trades and high returns for investment in business.

II. LITERATURE SURVEY

In the past , various approaches have been applied to analyze stock trend. Radu Iacomin [5] analyzed stock with the help of SVM .The shortcoming of the approach was discussed and moving forward, two new techniques were then evolved namely, GASVM and PCASVM.

Priti Saxena and Bhaskar Pant along with other fellow authors [6] ,generated the linguistic forms for raw data comprising of opening and closing price. Apriori model was then applied on the data to find out the most frequent patterns.

Prof. Dr. K. Raza[7] analyzed stock with the help of four different versions of Artificial Neural Network. It went on to establish an association of stock behavior with change in petrol prices and finally concluded that it was possible to analyze stock with the help of machine learning techniques.

In our model, we first applied different existing algorithms on stock market and compared their results. Then in our proposed model, we provided a model which would check itself at each stage for the correctness in analyzing share trends.

III. IMPLEMENTATION

A. Black-Hole Algorithm

We applied the black hole algorithm for analyzing stock trend. For the same, we first calculated data points for each factor of the share i.e.(high price, low price , opening price and closing price) in such a way that the data points lie in the range of that factor (i.e. in between the maximum value and the minimum value of that share factor).We then moved ahead and calculated the fitness value for each black hole star through a series of iterations by comparing the distance of a particular star from the black hole .The one which has minimum distance corresponds to the star that is closest to many data points and thus would be assigned the position of black hole. We have also put up a condition to check whether a star has submerged into black hole or not. If it has, then we replace that star with a new star(data point).[1]We then move a star nearer to black hole. The whole process is repeated for further days' analysis whilst taking into account all the previous days' values as well .Black hole Algorithm is applied over a set of NSE data of one year , we get the following Accuracy graph for four factors under consideration Open Price, Close Price, High Price, Low Price.

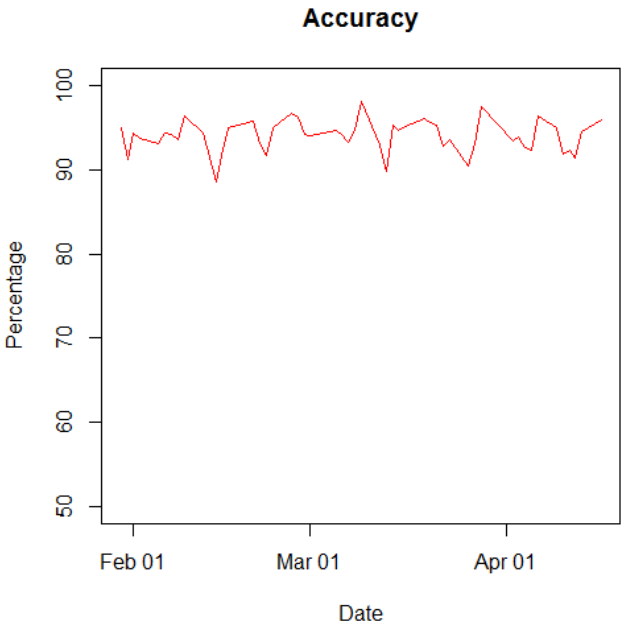

Fig: 1

B. Particale Swam Optimization(PSO)

We applied the PSO technique for stock analysis. It is to be noted that PSO draws a lot of parallels with black hole. One notable difference however is the calculation of personal best for each data point. For the first iteration it is taken same as the data point but moving forward it is calculated whilst comparing the fitness values. Also velocity is introduced and it is calculated with reference to personal best and global best of a data point where global best has the same meaning as the centroid had in the black hole algorithm. Thus , we analyzed the accuracy with which it offers the results for various factors .[1]Particle Swam Optimization is applied over a set of NSE data of one year , we get the following Accuracy graph for four factors under consideration Open Price, Close Price, High Price, Low Price.

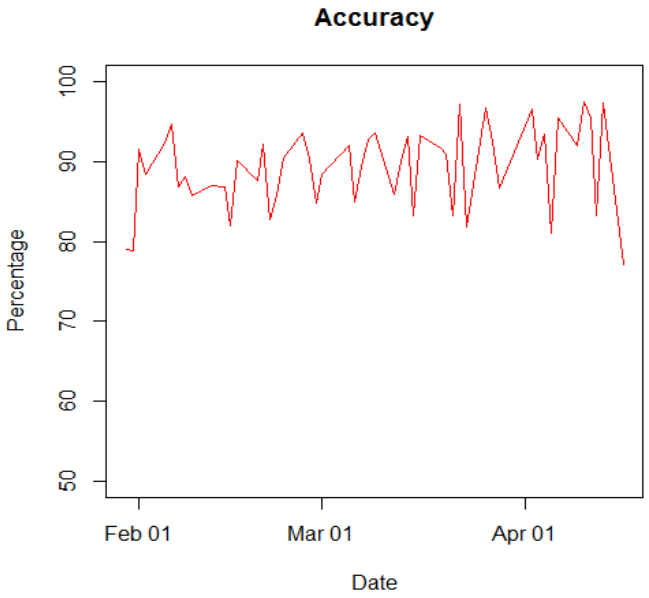

Fig: 2

C. Decision Tree

Here we analyzed the value of closing price of a share corresponding to other factors such as high price, opening price, low price and volume. For the same , we first prepared a matrix comprising of zeroes and ones where a zero indicates that the value of share price has fallen from the previous day whilst moving towards the next one and a one indicates that it has risen in comparison with the previous days' value. [4]
We then made closing price a function of other factors and analyze the probability of whether its value would fall or rise (over the testing set) and how other factors affect the probability.

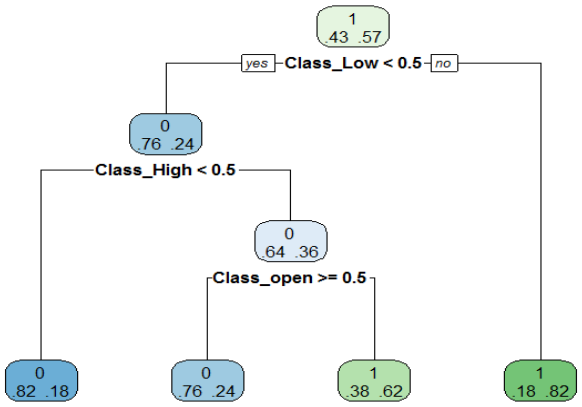

Fig: 3

D. Naive Bayes Algorithm

We analyzed the value of closing price of a share corresponding to other factors such as high price ,opening price ,low price and volume. We therefore prepared a matrix comprising of ones and zeroes where their meaning is same as it was in decision tree. We then bind the various factors together for better analysis. We made closing price a function of other factors and tested our data[2] .It is to be noted that the probability would come with thirty two as the denominator since we are taking five factors into account. We at last prepared a confusion matrix to test the accuracy of our results. Its accuracy is calculated as 85.41% over a set of NSE data of one year.

IV. Proposed Model

In the proposed model , we divide the dataset into two parts i.e. the training set and the testing set. The training set is used to train the data and learn about the share qualities while the testing set is used for prediction.

In the training set, we calculate the absolute value of change in a share price over the time interval for every day and take the mean of it. This gives us the average value by which a share price can rise or fall .Also we prepare a matrix wherein if the value of share price increases on a particular day with respect to the previous one we mark that field in the matrix as 1 otherwise if the share price falls then the field is marked as 0.Also each row of the matrix corresponds to the day number.

Now we count the number of 1s and 0s in the prepared matrix .If the number of 1s are more than the number of 0s then we classify the share as a good share otherwise it would be designated as a bad share. This step is important for the prediction of share price for the first day in the testing set. we prepare a prediction matrix and would assign it the values 1 or 0 according to our prediction subsequently. Let its name be y. If the share is a good share then the value of the share price is increased by the change we calculated earlier and the first row of the newly created prediction matrix is marked as 1 otherwise if the share is a bad share then the value is decreased and correspondingly the first row of the prediction matrix is marked as 0.Now we again prepare a similar matrix for testing set as well in terms of 1s and 0s on the basis of actual rise and fall in the price of share. Let its name be x .Now we move on to the next set of days for prediction and would repeat the following steps for every prediction henceforth. We check here whether our first day's prediction was in the right direction by comparing the values of first row of x and y matrix. If there is a match ,that means our prediction is in the right direction and we further proceed for the next day accordingly by what we did for the previous one(i.e. if we increased the share price in the previous prediction then for the next prediction as well ,we do the same and vice versa) . The next field of the y matrix is marked accordingly. If , however the values of x and y do not match ,we rectify our error by changing the share price value for the next prediction by a factor of twice the change we calculated earlier.(So for example if the share price falls and we predicted a rise so for the next prediction we now alter the predicted value of the previous day by a factor of twice the change and decrease its(predicted price) value accordingly). Also we mark the next field of y matrix as opposite to the previous one. In this way, we calculate the predicted price and subsequently the accuracy of prediction by comparing it to actual values.

V. ALGORITHM: PROPOSED MODEL

Input: A Particular company's dataset comprising of the different values or factors for that particular day

Output: accuracy with which datset was analyzed

1. Read the dataset and partition it into two subsets namely the training set *n1* and the testing set n_2

2.Prepare a matrix *z* for storing ones and zeroes having number of rows equal to size of training set and columns equal to number of factors included for the share

3. Initialize change *c* for calculating change in share value.

4. $I = 0$

5. **While** each $I < n1$ **do**

6. Calculate *c* based on absolute change in share price over $I-1$ and I days

7. Set I *th* row as one or zero in matrix *z* depending on increase or decrease in share value

8. **End for**

9. Count number of ones and zeroes in *z*

10. Initialize a new matrix *y* for storing predicted trends

11. *sign* <- 1

12. **if** number of ones >= number of zeroes then

13. Prediction for *(n1+1) th* day=value at *n1th* day + *c*

14. Update first row of *y* with ones

15. *sign*=1

16. **Else**

17. Prediction for *(n1+1)th* day=value at *n1th* day-*c*

18. Update first row of *y* with zeroes

19. *sign* = -1

20. **End**

21. Prepare a matrix x for storing actual trend in share price in the form of ones and zeroes

22. **For each** *I* < *n2* do

23. Set *I th* row as one or zero in matrix *x* depending on increase or decrease in share value

24. **End for**

25. *J = n1+2*

26. *H=1*

27. **For each** *J* < *n2*

28. **if** *H th* row of *x*== *H th* row of *y*

29. Prediction for *Jth* day = Prediction of *(J-1)th* day+(sign)*c

30. Update *(H+1)th* row of *y* with the same values as in *(H) th* row

31. **Else**

32. *sign=sign**-1

33Prediction or *J th* day=prediction of *(J-1)th* day+2*(sign)*c*

34. Update *(H+1)th* row of *y* with the opposite values as in *(H)th* row

35. **End if**

36. *H++*

37. **End for**

38. Calculate accuracy *a* for each predicted value by comparing to actual values

---

VI. TEST CASES ON PROPOSED MODEL

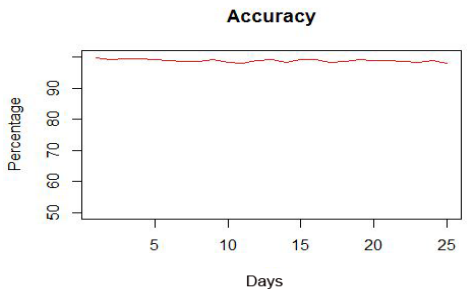

Fig4: 25 days of testing data, 222 days of training

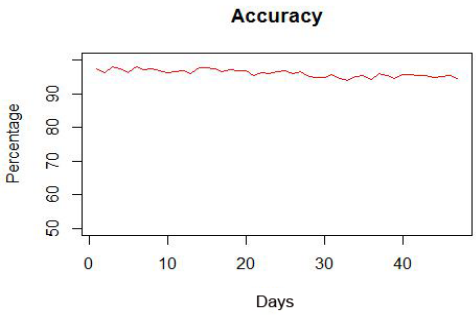

Fig5: 47 days of testing data, 200 days of training

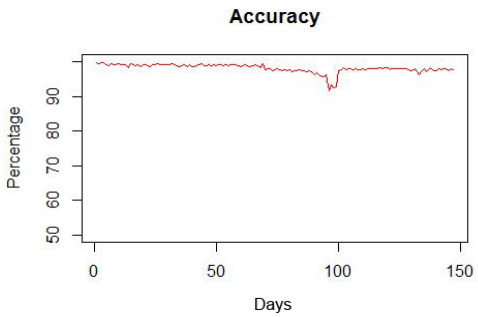

Fig6: 147 days of testing data, 100 days of training

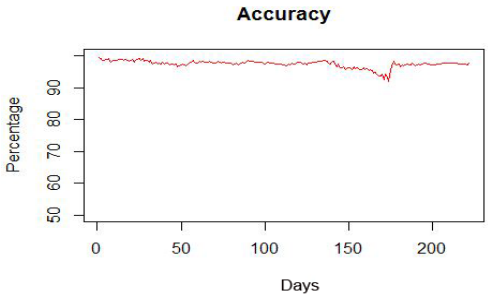

Fig7: 222 days of testing data, 25 days of training

It can be inferred from Fig4-5 that as we increase the number of days of training data i.e. the training set , the graph assumes a constant shape. On the other hand as we can see in Fig6-7 decrease in the number of training days i.e. the training set led the graph to have depressions at some points and projections at others.

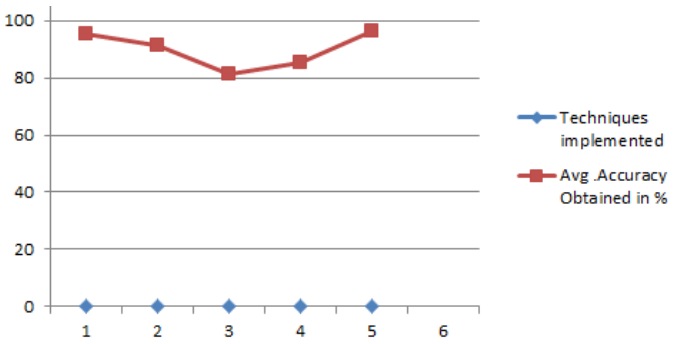

Fig8: Comparison of mean of accuracy of various techniques

S. No.	1	2	3	4	5
Algorithm	Black-Hole	PSO	Decision Tree	Naive Bayes	Proposed Model
Mean Accuracy	95.10	83.57	81.16	85.56	96.60

Table:1

VII. CONCLUSION

Evaluating the Stock market has at all times been a tough work for analysts. Thus, we attempt to make use of large data sets to forecast the stock market indices.. As already stated earlier market trends depend on various external parameters. Factors such as political scenario, currency exchange rate, economy of the state etc are not in one's hands .In our work we compared the performance of different predefined techniques by testing them on the same dataset. We also applied a few changes in some of the factors of the share whilst analyzing that share in order to accommodate those techniques in our approach.

We then proposed our model wherein we took the same dataset and tested it on a number of conditions. The results were obtained and the accuracy with which our model worked for that dataset were put up .Though our model is a simplified one, it can be further improvised and many a concept can be built upon that framework to further increase its productivity and its capability to endure drastic changes in the market condition.

Thus, it would be suffice to say that there is a lot of room to develop more effective techniques to take the above stated factors into account. Also if there is a sudden crash or boom in the market, there is a need for developing efficient techniques to forecast as to which trend would the market follow in the future. Such unpredictability is what makes the study of share trends so interesting . We have given our best efforts for the same to reduce the intricacies of stock market by analyzing it in our capacity.

We believe that there is a lot of scope for improvement in forecasting techniques for share market and we hope that our work is just a beginning in employing different models to analyze the market. Financial analysts / Academic researchers can use such analyzing models and develop hybrid ones to take trading decision by observing market behavior.

References

[**1**] Farahmandian M, Hatamlou A (2015) Solving optimization problems using black hole algorithm. J Adv Comput Sci Technol 4(1):68–74

[**2**] B. Liu, E. Blasch, Y. Chen, D.Shen, G. Chen, "Scalable Sentiment Classification for Big Data Analysis Using Naive Bayes Classifier," IEEE Intl Conf. on Big Data, Oct. 2013.

[**3**] C.W. Tsai, C. H. Hsieh, and M.C. Chiang, "Parallel Black Hole Clustering Based on MapReduce" , *IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2015*.

[**4**] A. Navada, et al., "Overview of use of decision tree algorithms in machine learning," in Control and System Graduate Research Colloquium (ICSGRC), 2011 IEEE, 2011, pp. 37-42.

[**5**]Radu Iacomin (2015),"Stock Market Prediction",2015 19th International Conference on System Theory, Control and Computing (ICSTCC), October 14-16, Cheile Gradistei, Romania,Pages:200-205,DOI:10.1109/ICSTCC.2015.732 129 3.

[**6**] Saxena, P., Pant, B., Goudar, R.H., Srivastav, S., Garg, V., Pareek, S.: Future predictions in Indian stock market through linguistic-temporal approach. In: 2013 7th International Conference on Intelligent Systems and Control (ISCO) (2013)

[**7**]Prof. Dr. Kamran Raza Dean FEST, Iqra University Karachi, Pakistan,"*Prediction of Stock Market by using Machine Learning Technique*".