Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Science and Engineering

Elixir Comp. Sci. & Engg. 70 (2014) 24058-24062

ISSN: 2229-712X

Knowledge Discovery in a Stock Data using Moving Averages

Mukesh Kumar* and Arvind Kalia

Department of Computer Sciences, Himachal Pradesh University, Shimla, India.

ARTICLE INFO

Article history:

Received: 5 March 2014;

Received in revised form:

25 April 2014;

Accepted: 5 May 2014;

Keywords

Stock data mining,

Majority voting,

Moving Average,

Exponential Moving Averages.

ABSTRACT

Association rules mining algorithms can be used to discover all item associations (or rules) in a dataset. Majority voting is adopted as classification technique and on the basis of voting pattern, the consequent is chosen. The moving averaging is applied on the obtained consequents to identify the emerging pattern. Four moving averages on the basis of Fibonacci sequence are applied. It has been observed that number of trades is more in lower range of moving averages as compared to higher range and a longer days averaging has not been yielding good returns. It has been observed that the accuracy level is higher in case of smaller duration whereas the error rate is on the higher side in case of longer period averages.

Introduction

Stock data mining has a potential to provide information for trading decision support and market surveillance [5,11]. The stock data mining is providing an ease for finding the relationships between stocks and their respective pattern[9]. For association rule mining, the target of discovery is not predetermined, while for classification or prediction there was one and only pre determined target. These two approaches have been integrated in different tools or techniques for classification based association rule discovery or called associative classification. Best confidence, majority voting and sliding window were used for associative classification in [15] and it has been established that majority voting found to be best among three. By converting the numerical stock time series data into symbolic sequences [7,14] , the association rules could be generated for a kind of market observations for a selected period. These tools in [1,8] in a first phase, typically discover all association rules and then post-process the resulting rule sets to retain only a small number of suitable rules. Several studies [1, 2, 6] have shown that the obtained rule sets or consequents often perform better or comparable to those obtained using more traditional rule learning algorithms such as in [3,4] .

Research Methodology

For the purpose of this study, the stocks dataset of eight years period i.e. from Jan. 2001 to Dec.2008 of NSE stock exchange is used [16]. The trading of the stock market within a day is recorded in a single text file. Each line represents the trading information of a stock. The fifty stocks of Nifty is chosen for study, but the complete data of only 25 stocks is made available so the 25 stocks were selected for this purpose. A list of stocks selected for data mining is presented in Table 1.

A novel approach has been presented for associative classification that used two types of strategies. It is aimed at finding a qualified consequent using majority voting approach. The second approach is aimed to find a timely reversal in an existing trend so that a trade is being initiated. The second approach is based on moving average. Majority voting is adopted as an associative classification technique and on the basis of voting pattern, the consequent is chosen.

Table 1 . List of stocks

S.no.	Code	Name
1	S1	ABB
2	S2	ACC
3	S3	BHEL
4	S4	BPCL
5	S5	CIPLA
6	S6	GRASIM
7	S7	HDFC
8	S8	HDFC BANK
9	S9	HLL
10	S10	ITC
11	S11	INFOSYS
12	S12	L&T
13	S13	M&M
14	S14	ONGC
15	S15	RANBAXY
16	S16	RELCAP
17	S17	RELIANCE
18	S18	SAIL
19	S19	SBI
20	S20	SIEMENS
21	S21	SUNPHARMA
22	S22	TATAMOTOR
23	S23	TATASTEEL
24	S24	UNITECH
25	S25	WIPRO

After that intra stock mining is conducted on that consequent for the next year and its trend is to be determined. Intra-stock pattern mining concerns with the discovery of repetitive temporal association patterns for the stock itself across a time span of few trading days [11]. Technical analysis tools are discussed and Moving Average technique is adopted to decide what type of action is required in case of reversal in the existing pattern. A combination of these two techniques is presented and the obtained results are discussed. The NSE data of eight years from year 2001-2008 is used and 25 stocks of different sectors have been chosen for the study.

Majority Voting

Majority Voting [15] is used to determine the effectiveness of the mined rules and on the basis of this approach the consequent is chosen and its accuracy is determined. The

Tele:

E-mail addresses: mukeshk.gzb@gmail.com

occurrences(voting] of each consequent is computed using the instances of the same and then they are categorized as per their votes and the consequent obtaining highest votes is declared as the selected consequent on which intra stock mining is to be conducted. The thirteen tables of association rules are generated. There are ten rules in table 1 and according to the majority approach the classification result should be "S5". It is because total votes for "S5" is 84 + 119 + 88 + 111 + 92 +86 = 580, total votes for "S8 " is 84 +114= 198 and total votes for S21 is 84 = 91 = 175. So as per majority approach S5 would be treated as winning consequent. Further this approach need not to be adopted if there is clearly shown highest number of occurrences of the same consequent in different rules, In case of a tie the total votes are required to be counted. Because of limited space one table is shown with all association rules with a minimum specified confidence. Maximum 10 association rules are taken for the study with minimum support value. In some cases more rules are generated while in some cases, few rules are generated but for the sake of clarity, maximum 10 rules are considered.

Table 2. Single consequent with majority voting for year 2001 support = .16

Rules	Association Rules	Confidence	Instances
1	s12,S25 -> S16	.88	43
2	S6,S11 -> S16	.88	42
3	S11,S16 -> S25	.84	49
4	S6,S25 -> S16	.83	49
5	S6,S16 -> S25	.81	44
6	S12,S16 -> S25		43
7	S23,S25 -> S16	.80	44
8	S16,S23 -> S25	.80	45
9	S6,S25-> S11	.75	42
10	S17 -> S16	.73	49

Table 3 Voting pattern of consequents 2001

Consequent	Votes
S11	1
S16	5
S25	4

With the use of Majority Voting the voting pattern of the consequents is presented in Table3 to Table 10. On that basis such eight tables having the list of consequent obtained using majority voting are generated. All the consequents with their voting pattern are presented in the separate tables for each year.

Table 4. Voting pattern of consequents for the year 2002

Consequent	Votes
S5	6
S8	2
S21	2

Table 5. Voting pattern of consequents for the year 2003

Consequent	Votes
S18	5
S22	1
S23	4

Table 6. Voting pattern of consequents for the year 2004

Consequent	Votes
S10	3
S18	3
S22	2
S23	2

Table 7. Voting pattern of consequents for the year 2005

Consequent	Votes
S6	6
S10	2
S21	2

Table 8 Voting pattern of consequents for the year 2006

Consequent	Votes
S1	2
S16	1
S18	2
S20	1
S21	2
S23	3

Table 9 Voting pattern of consequents for the year 2007

Consequent	Votes
S3	2
S12	1
S16	3
S18	4

Table 10. Voting pattern of consequents for the year 2008

Consequent	Votes
S3	1
S12	3
S16	4
S17	1
S24	4

A consolidated table is prepared of all the eight years on the basis of majority voting and the consequent was decided and presented in Table 11.

Table 11 the consequent with majority voting is presented below in a table

Year	Consequent	Total Instances	Support
2001	S16	223	.16
2002	S5	578	.33
2003	S18	186	.18
2004	S18	134	.18
2005	S6	265	.17
2006	S23	123	.16
2007	S18	164	.17
2008	S24	164	.33

Moving Averages

A moving average is a trend following device. Its purpose is to identify that a new trend has begun or that an old trend has ended or reversed [13]. Its purpose is to track the progress of the trend. The moving average is a follower, not a leader.. The moving average follows a market and conveys that a trend has begun, but only after the fact. Moving averages is one of the most popular and easy to use indicators available to the technical analyst [10, 12]. MA method enable the construction of a computerized algorithm for the application of the method, and the indications of buy or sells signals. Because past price data is used to form moving averages, they are considered lagging, or trend following, indicators. A moving average has been an average of observations from several consecutive time periods. To compute a moving average sequence, we compute successive averages of a given number of consecutive observations. The objective underlying the MA method is to smooth out seasonal variation in the data. The method involved a comparison of the most recent market price with the long MA of the price. If the current price is higher than the long MA, a long position (BUY) should be adopted, and conversely, if the current price is lower than the MA, a short position (SELL) could be adopted.

Exponential Moving Average (EMA)

Exponential moving averages (also called exponentially weighted moving averages) apply weighting factors which decrease exponentially. EMAs reduce the lag by applying more weight to recent prices relative to older prices. The shorter the EMAs period, the more weight would be applied to the most recent price. Shorter moving averages would be more sensitive and generate more signals. However, there would also be an increase in the number of false signals and whipsaws. Longer moving averages would move slower and generate fewer signals. These signals would likely prove to be more reliable, but they also might come late. Because moving averages followed the trend, they work best when a security is trending and are ineffective when a security move in a trading range. A simple visual assessment of the price chart could determine what a security exhibited. Now different Fibonacci days strategies EMAs was studied in the further study.

Characteristics Of Trend

There were three ways to identify the direction of the trend with moving averages: direction, location and crossovers. The first trend identification technique used the direction of the moving average to determine the trend [9]. If the moving average was rising, the trend was considered UP. If the moving average was declining, the trend was considered DOWN. The direction of a moving average could be determined simply by looking at a plot of the moving average or by applying an indicator to the moving average. In either case, we would not want to act on every subtle change, but rather look at general directional movement and changes.

Experimental Results

Four combination of Fibonacci days are used as ema8-21, ema13-34, ema 21-55 and ema 34-89. The following charts have been presented using these four different Moving Average strategies as ema 8-21, ema13-34, ema 21-55 and ema 34-89 for each consequent for each year and the crossover details of that consequent was tabulated.

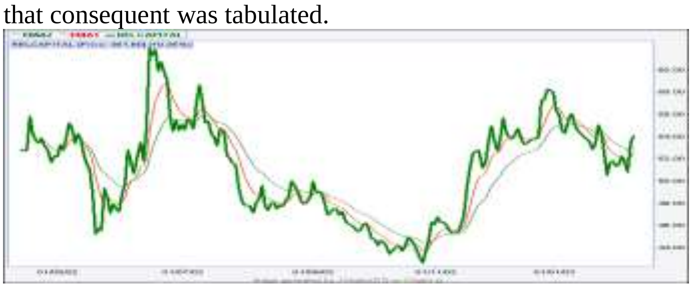

Fig 1 s16 chart with ema8-21

Fig. 2 s16 chart with ema13-34

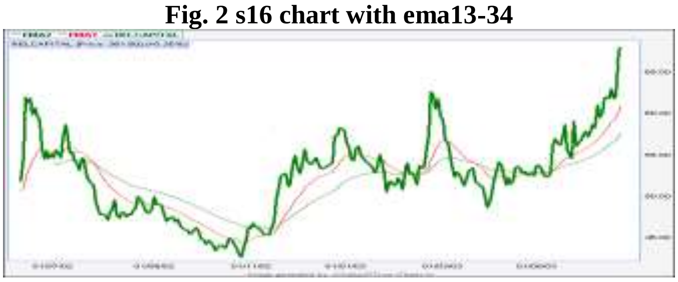

Fig. 3 s16 chart with ema21-55

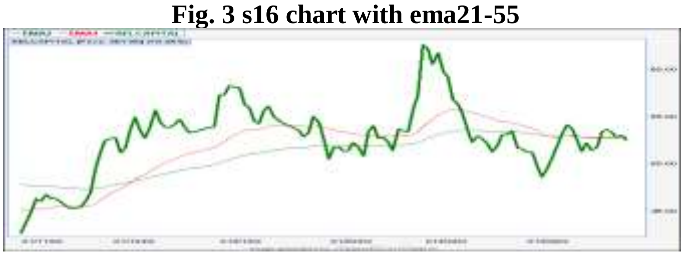

Fig. 4 s16 chart with ema34-89

Table 12: Crossover details of stock s16 for year 2002

ema	date	buy	date	sell
8-21	6.6.2002	51.80	15.5.2002	52.35
	6.6.2002	51.80	15.7.2002	54.25
	1.11.2002	46	15..7.2002	54.25
	1.11.2002	46	31.12.2002	58.15
13-34	12.6.2002	51.80	19.7.2002	54.40
	15.11.2002	46	19.7.2002	54.40
	15.11.2002	46	31.12.2002	58.15
21-55	20.11.2002	49.45	25.7.2002	51.45
	20.11.2002	49.45	31.12.2002	58.15
34-89	29.11. 2002	53.60	31.12.2002	58.15

On the same pattern charts for all the six years Charts were used to determine the crossover of emas and accordingly six tables are presented below

Table 13. crossover details of stock s5 for year 2003

ema	date	buy	date	sell
8-21	30.5.2003	55	20.1.2003	69
	30.5.2003	55	31.12.2003	105
13-34	4.6.2003	57	12.1.2003	69
	4.6.2003	57	31.12.2003	105
21-55	16.6.2003	59	20.1.2003	69
	16.6.2003	59	30.12.2003	96
34-89	11.7.2003	66	30.12.2003	1-5

Table 14. crossover details of stock s18 for the year 2004

ema	date	buy	date	sell
8-21	20.4.2004	37	21.1.2004	45
	20.4.2004	37	30.4.2004	35
	6.7.2004	30	30.4.2004	35
	6.7.2004	30	30.12.2004	63
13-34	14.7.2004	30	20.1.2004	46
	14.7.2004	30	31.12.2004	63
21-55	30.7.2004	40	13.2.2004	47
	30.7.2004	40	31.12.2004	63
34-89	17.8.2004	42	15.3.2004	37
	17.8.2004	42	31.12.2004	63

Table 15 crossover details of stock s18 for the year 2005

ema	date	buy	date	sell
8-21	18.1.2005	61	12.1.2005	56
	18.1.2005	61	28.3.2005	64
	12.7.2005	52	28.3.2005	64
	12.7.2005	52	10.10.2005	62
	30.12.2005	53	10.10.2005	62
13-34	20.7.2005	54	8.4.2005	64
	20.7.2005	54	13.10.2005	57
	31.12.2005	55	13.10.2005	57
21-55	3.8.2005	54	18.4.2005	54
	3.8.2005	54	19.10.2005	55
	31.12.2005	55	19.10.2005	55
34-89	9.8.2005	62	2.5.2005	54
	9.8.2005	62	25.10.2005	53
	31.12.2005	55	25.10.2005	53

Table16 crossover details of stock s6 for the year 2006

EMA	date	buy	date	sell
8-21	21.1.2006	1375	15.5.2006	2125
	29.6.2006	1835	15.5.2006	2125
	29.6.2006	1835	29.12.2006	2997
13-34	21.1.2006	1375	19.5.2006	1925
	9.7.2006	1901	19.5.2006	1925
	9.7.2006	1901	29.12.2006	2997
21-55	24.7.2006	1967	30.5.2006	1876
	24.7.2006	1967	29.12.2006	2797
34-89	12.7.2006	2004	13.6.2006	1621
	12.7.2006	2004	29.12.2006	2997

Table 17. Crossover details of stock s23 for the year 2007

ema	date	buy	date	sell
8-21	15.1.2007	486	5.1.2007	469
	15.1.2007	486	2.2.2007	470
	4.4.2007	438	2.2.2007	470
	4.4.2007	438	1.8.2007	624
	30.8.2007	657	1.8.2007	624
	30.8.2007	657	9.11.2007	834
	20.12.2007	825	9.11.2007	834
13-34	5.4.2007	465	6.2.2007	464
	5.4.2007	465	14.8.2007	641
	3.9.2007	691	14.8.2007	641
	3.9.2007	691	31.12.2007	935
21-55	10.4.2007	495	21.8.2007	546
	31.8.2007	690	21.8.2007	546
	31.8.2007	690	31.12.2007	935
34-89	17.4.2007	528	31.12.2007	935

Table18. Crossover details of stock s18 for the year 2008

ema	date	buy	date	sell
8-21	21.2.2008	236	8.1.2008	261
	21.2.2008	236	12.3.2008	218
	1.8.2008	144	10.9.2008	139
	30.12.2008	79	10.9.2008	139
13-34	27.2.2008	243	11.1.2008	250
	27.2.2008	243	12.3.2008	218
	8.8.2008	146	12.3.2008	218
	8.8.2008	146	12.9.2008	140
	31.12.2008	78	12.9.2008	140
21-55	30.12.2008	79	22.1.2008	203
34-89	4.3.2008	229	1.2.2008	185
	4.3.2008	229	31.12.2008	78

Four moving averages as ema8-21, ema13-34, ema21-55 and ema34-89 are applied. It has been observed that number of trades is more in lower range of moving average as compared to higher range. Now evaluating the patterns for different stocks using different emas, return is being evaluated as once buy call is initiated, it would be terminated after another SELL signal and then it has to wait for another BUY signal. Consolidated returns with all different EMAs are presented in Table 19.

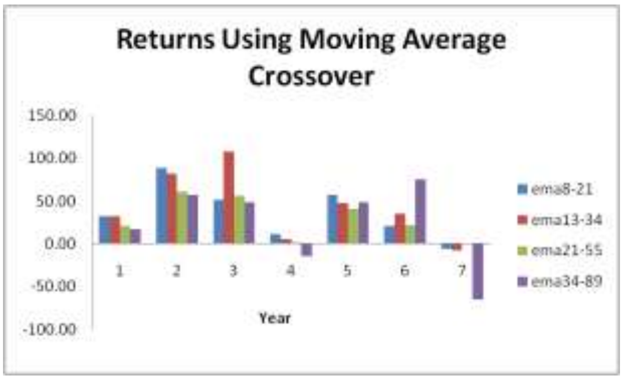

Fig. 5 Chart of four emas

Table 19. Moving Averages returns

year	ema8-21	ema13-34	ema21-55	ema34-89
2002	32.93	32.93	21.21	16.92
2003	89.09	82.53	61.46	57.91
2004	51.25	107.80	56.35	49.00
2005	11.83	5.44	1.81	-14.23
2006	57.76	47.85	41.35	48.56
2007	21.60	35.85	22.45	75.54
2008	-5.44	-7.05	0.00	-64.62

In this approach the return is found to be positive for all the years except in one in which it is found to be on a negative side, but still it is on a much lower side. Further with ema34-89 the return is having negative value for two years Fig. 5, but those values found to be on a higher side, thus indicating, that perspective is to be avoided.

So a longer days averaging has not been yielding good returns. So in both the approaches the ema34-89 strategy is found to be not a better strategy and such a longer days averaging approach might be avoided.

The performance metric such as accuracy and error rate is defined as follows

Accuracy = Number of correct predictions / Total number of predictions

Error rate = Number of wrong predictions / Total number of predictions

The accuracy and error rate for each ema strategy and for Moving Averages approach was evaluated and presented below in table 20. .It has been observed the accuracy level was higher in case of smaller duration whereas the error rate is on the higher side in case of longer day average.

Table 20. EMA comparison

	ema8-21	ema13-34	ema21-55	ema34-89
accuracy rate	.63	.8	1	.71
error rate	.37	.2	0	.29

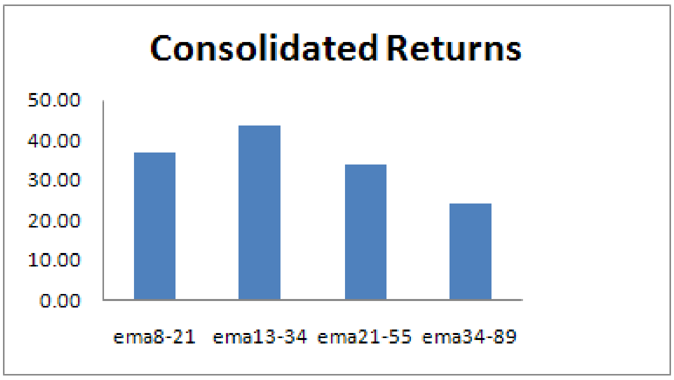

Fig. 6 Returns using four emas (Moving Averaging)

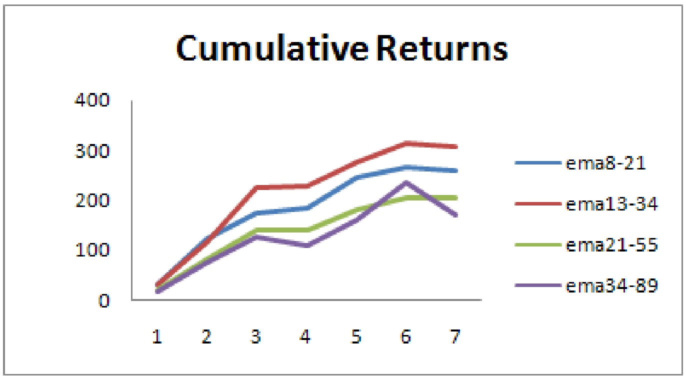

Fig 7. Cumulative returns of different ema strategies

It has been shown that except in one year there was an overall positive return for each year. There is a negative return in one year but was still on much lower side. The returns is evaluated using all the four averages and it has been found that overall positive returns is obtained for different days moving strategies Fig. 6. There are variations in the terms of returns obtained and the highest return is achieved with ema13-34. It is also clearly reflected in the cumulative returns chart in fig. 7 which indicates that using ema13-34, the highest profitability was achieved.

Conclusion

It has been established that using majority approach on the generated rules one could plan their investment strategies, but it has not been giving a positive outlook for all the years, therefore only majority voting approach is not sufficient, some more indicators are required to be applied after having obtained a consequent using majority voting. Moving averages is a technical indicator which normally used to spot a trend. Normally it is very useful indicator when there is a change in the market trend either uptrend or downtrend, It is used to have a confirmation of a reversal of the trend. Moving averages can be an effective tool to identify and confirm the trend, identify support and resistance levels, and develop trading systems. However, traders and investors should learn to identify securities that are suitable for analysis with moving averages and how this analysis should be applied. Usually, an assessment can be made with a visual examination of the price chart, but sometimes it will require a more detailed approach. Four moving averages as ema8-21, ema13-34, ema21-55 and ema34-89 were applied. It has been observed that number of trades was more in lower range of moving average as compared to higher range. the return was found to be positive for all the years except in one year it found to be on a negative side, but still it is on a much lower side. Further with ema34-89 the return is having negative value for two years, but those values found to be on a higher side, thus indicating that perspective found to be avoided. Further a longer days averaging has not been yielding good returns. So in both the approaches the ema34-89 strategy is found to be not a better strategy and such a longer days averaging approach might be avoided. It has been observed the accuracy level is higher in case of smaller duration whereas the error rate is on the higher side in case of longer day average. Therefore it is established that using ema13-34, the highest profitability is achieved among the four emas.

References

1. Albrecht Z. and Luc D. R., “ CorClass: Correlated Association Rule Mining for Classification”, Springer-Verlag Berlin Heidelberg, pp. 60-72, 2004.
2. Agrawal R., Srikant R., “Fast Algorithms for Mining Association Rules”, Proceedings of the 20th VLDB Conference Santiago, Chile 1994.
3. Agrawal, R., Imielinski T. and Swami A., “ Mining Association Rules between Sets of Items in Large Databases” Proceedings of the ACM SIGMOD Conference ,New York, USA, pp207-216,May 1993.
4. Armano, A. Murru and Roli F... “Stock Market Prediction by a Mixture of Genetic Neural Experts.” International Journal of Pattern Recognition and Artificial Intelligence Vol. 16, No. 5, pp 501-526. 2002.
5. Gencay, R. and Stengos, T. ,” Moving average rules, volume and the predictability of security returns with feed forward networks”, Journal of Forecasting, 17(5/6),pp 401–414, 1998.
6. Han, J. and Kamber, M. “Data Mining Concepts and Techniques”. Morgan Kauffman Publishers, Elsevier, San Francisco, CA.2006.
7. HsiehY.S., Yang D.L., and Jungpin W. “Using Data Mining to Study Upstream and Downstream Causal Relationship in Stock Market”, JCIS, 2006.
8. Last M, Klein Y and Kandel A. “Knowledge Discovery in Time Series Database.” IEEE Transactions on Systems, Man, and Cybernetics-part b, VoL 31(1).2001.
9. Leleu1,M,, Boulicaut J.F., “Signing stock market situations by means of characteristic sequential patterns”, In Proc. 3rd International Conference on Data Mining and Database for Engineering, Bologna, pp 655-664, 2002.
10. Murphy J.J., “Technical Analysis of the Financial Markets”, New York Institute of Finance, U.S.A. 1999.
11. Nayak R.and Paul B.T., “ Temporal Pattern Matching for the Prediction of Stock Prices” In Second Workshop on Integrating AI and Data Mining (AIDM).2007.
12. Oberlechner, T. “ Importance of technical and fundamental analysis in the European foreign exchange market”, International Journal of Financial Economics, 6(1),pp 81–93.2001.
13. Ou J, Penman S.” Financial statement analysis and Prediction of stock returns”, Journal of Accounting and Economics 11:pp295-329. 1989.
14. Soon, L.-K. and Lee, S.H. , “ Explorative Data Mining on Stock Data Experimental Results and Findings” . In Proceedings of the 3rd International Conference on Advanced Data Mining and Applications, Harbin, China , pp562-569. 2007.
15. Ting, J., Fu, T. and Chung, F. “Mining of Stock Data: Intra- and Inter-Stock Pattern Associative Classification”. In Proceedings of 2006 International Conference on Data Mining, Las Vegas, USA, pp30-36.,2006.
16. http : www.nseindia.com