

A Study on Stock Market Analysis for Stock Selection – Naïve Investors' Perspective using Data Mining Technique

B. Uma Devi
Assistant Professor
R.D.Govt. Arts College
Sivagangai

D. Sundar
Assistant professor
Thiagarajar School of Mgt.
Madurai

Dr.P. Alli
Prof &Head (CSE)
Velammal College of Engg&Tech.
Madurai

ABSTRACT

An insight of stock market trends has been an area of vast interest both for those who wish to make profit by trading stocks in the stock market. Generally there is an opinion about stock market like high risk and high returns. Even though we have a huge number of potential investors, only very few of them are invested in the stock market. The main reason is the inability of risk taking skill of investors. Though get low returns they want to save their money. One important reason for this problem is that, they don't have a proper guidance for making their portfolio. In this paper we focus the real world problem; we had selected three indices such as CNX Realty, BANK NIFTY and MIDCAP 50. The analysis is purely based on the data collected from past three years. The Data mining technique, Time series interpretation is applied for the Data analysis to show the ups and downs of a particular index. The correlation and Beta are the tools which gives the suggestion about the stock and its risk. The correlation tool is used to identify the relationship between the index and company individually. This Beta is used to identify the risk associated with the stock.

Categories and Subject Descriptors

[Data mining] Stock market analysis

General Terms

Data mining, Time series Analysis

Key Words

Index, Correlation, Beta, Time series

1. INTRODUCTION

The main components of financial market are money and capital markets. The Security market is divided into two categories, primary market and secondary market. The primary market is that part of the capital markets that deals with the issuance of new securities. The process of selling new issues to investors is called underwriting. In the case of a new stock issue, this sale is called as initial public offering (IPO).

The secondary market is an on-going market, which is equipped and organized with a place, facilities and other resources required for trading securities after their initial offering. It refers to a specific place where securities transaction among many and unspecified persons are carried out through intermediation of the securities firms, i.e., a licensed broker, and the exchanges, a specialized

trading organization, in accordance with the rules and regulations are established by the exchanges.

Stock market forecasting includes uncovering market trends, planning investment strategies, identifying the best time to purchase the stocks and what stocks to purchase. There has been a critical need for automated approaches to effective and efficient utilization of massive amount of financial data to support companies and individuals in strategic planning and investment decision making. It is essential and need to find need of the hour in the field of Market behavior is to reveal the market trends , investment decisions , investment strategies ,identifying the best time to purchase stocks and what stocks to purchase.

The stock market is a complex, non stationary, chaotic [1][13] and non-linear dynamic system. Forecasting stock market, currency exchange rate, bankruptcies, understanding and managing financial risk, trading futures, credit rating, loan management, bank customer profiling, and money laundering analyses are the core challenging tasks to be considered.

The novel idea is to give an insight for the naïve investors. The results analysis and explanation will provide noteworthy thoughtful for all new investors. This paper introduces an algorithm, which includes the tools Correlation and Beta, to interpret the relationship and the risk associated with the stock.

2. BACKGROUND STUDY

The Analysis was made, based on the past historical data of BSE and NSE. The recent research , in association with data mining techniques for Time Series using the algorithms like "ARMA " and AR are more useful for the above said massive data analysis of stock market. But what decisions oriented result it will give in naïve investors' outlook.

The main intention is to make wakefulness for the new contestant for stock market and also to remove the fear about stock related issues. There are many data mining techniques, with its own algorithms will support for massive data exploration. [2]Data mining is a technique of discovering useful patterns in data that are hidden and unknown in normal position.[3] It comes from several fields like statistics, [4]database machine learning.

It is more necessary to, understand the behavior of the stock market. This can be great challenge for all stock

market investors. In perspective of naïve investors, there is no standard system or guidelines for them to understand. It is an approach where it shows the stock market flow [5] [6] by reading the data from past three years. The data mining technique helps to predict [3] the eventual or sudden falls met by the stock market. The traditional algorithm like ARMA is running with its own disadvantages. To identify patterns in time series data, The Correlation and Beta calculations are the two new modified approaches is introduced to find out the future forecast for the bank nifty .It will give thoughtful idea to foresee the future trend about the particular index in the forthcoming seasons.

3. PROBLEM DEFINITION

In this paper the objective is to meet out the general challenge, i.e., the goal is to improve the decision making power and wakefulness about the investment in the stock market from the naïve user's perspective. The naïve investors are having the problem of choosing the valuable

Table 1. NSE indices

S & P CNX Nifty	CNX MIDCAP
JUNIOR	NIFTYMIDCAP 50
CNX IT	CNX INFRA
BANK NIFTY	CNX REALTY
CNX 100	S&P CNX DEFTY
S&P CNX 500	

4. THE RESEARCH PROPOSAL

4.1 Time series Analysis

The definition Time series states that it is an ordered sequence of values of a variable at equally spaced time intervals [7]. It can be obtained from any system at determined time interval[10]. The daily price change of a market, Process and quality control, Economic Forecasting, Census Analysis, Stock Market Analysis may be considered as a Time series. For Example consider the equation (a)

$$X = \{x_i \quad i = 1 \dots N\} \quad \rightarrow (a)$$

In this equation i is the time index and N is the total number of observations. The important events are generated over a period of time. Consider the immediate changes such as fall and raise met by the stock market. It helps to investigate and foresee the proceedings. The Box Jenkins or Autoregressive Integrated moving average (ARIMA) is a traditional time series used to

4.2 The Proposed Algorithm

Correlation calculation:

Declaration
X = list of banks in bank nifty
Mc = Market cap
Initialize
A[i] = BankNiftyclosingprice
B[i] = Individual Bankclosingprice

stock. The reason for this concern is the lack of knowledge about the market and lack of education. It is very essentials to identify the stocks which are suitable for the investors' Expectations and identify the suitable stocks for short term investors and long term investors based on the stock market analysis [4]. This project gives an easiest way to identify the stocks which is suitable for the investors' expectations. It means that this research will give suitable stocks for short term investors and long term investors based on the data analysis and its wave pattern movement. And also it will give the details regarding risk associated with these stocks.

3.1 Objectives

- To give an overview of stock indices flow
- To give an insight about the changes happened in day to day market state.
- To give Guidance to the naïve investors initially, investors may free from the fears about the risk psychologically

Table 2. List of banks under Bank Nifty

Union bank of India	HDFC Bank
Axis Bank	ICICI Bank
Bank of Baroda	IDBI Bank
Bank of India	Kotak mahindra bank
Canara Bank	Oriental bank
State Bank of India	Punjab National Bank

model such time series. Nevertheless, the ARIMA method is limited by the requiremnt of stationary of time series and control of residuals.

The event depiction function changes according to the forecast plan. For Example x_i represents the today's closing price of stock and it is necessary to predict the percentage changes of tomorrow's price, the event depiction function can be defined in the following equation (b).

$$g(t) = \frac{x_{i+1} - x_i}{x_i} \quad \rightarrow (b)$$

The main drawback is that the time series should be converted into stationary [9] and periodic series to analyze it. As a promising discipline[22] data mining is the process of discovering hidden and useful information from enormous data. In this paper the proposed times series data mining [11]method is based on the insight made against the Data related to Bank nifty and rest of the Top five banks in the index.

Loop (1...n months)
Calculate
Avg = $\sum A[i] / \text{no. of working days};$
End;
Find the correlation {A [i] B [i]}
End;

Beta calculation:

Declaration

$$\beta = \frac{\text{Closing price (bank)} = r_p}{\text{Current closing price} = r_a} \times \frac{\text{Covar}(r_a, r_p)}{\text{Var}(r_p)}$$

5. DATA ANALYSIS AND INTERPRETATION

As mentioned early, there are three indices chosen for the study. From these each indices we had chosen five stocks which have more capitalization when compared to other stocks[23]. The following table and chart shows that M cap of the companies listed in the BANK NIFTY.

Table 3. Company Name vs. M Cap

Company Name	Mcap Rs(In Crores)
Kotak bank	34,765.14
Bank of India	17,729.28
IDBI	10,594.30
Bank of Baroda	30,434.14
Union Bank	15,123.32
Axis bank	46,708.44
Canara bank	19,217.34
PNB	31,112.93
Oriental bank	8,557.04
HDFC bank	113,029.00
SBI	123,542.43

Table 4. Nifty Month vs. Closing index Figure2. Nifty Month vs. closing point

Date	Bank Nifty/CNX Closing Point
Jan - Mar'09	4160.716667
Apr - Jun'09	6628.6
Jul - Sep'09	7929.5
Oct - Dec'09	8838.35
Jan - Mar'10	8944.633333
Apr - Jun'10	9566.216667
Jul - Sep'10	11091.23333
Oct - Dec'10	12024.93333
Jan - Mar'11	10927.55
Apr - Jun'11	11249.75
Jul - Sep'11	9965.116667

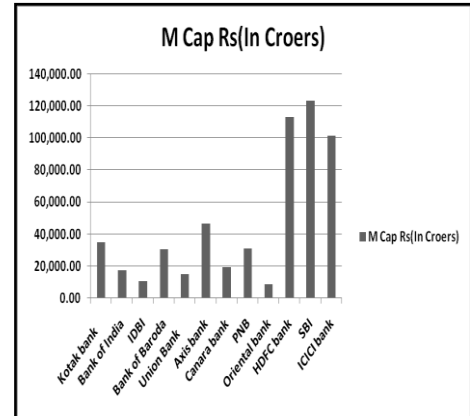
Table 5. Correlation between Bank Nifty/CNX with top 5 Banks

Date	BANK NIFTY/CNX Closing Point	SBIN Closing Price	HDFC Closing Price	ICICI Closing Price	AXIS Closing Price	KOTAK Closing Price
Jan - Mar'09	4160.716667	1088.290833	909.7766333	377.7294667	401.5554	280.1011333
Apr - Jun'09	6628.6	1500.755	1278.691667	590.2724667	645.7235667	530.8995333

If $\beta < 1$ then the stock has low risk
 Else
 Stock has high risk
 End if;
 End.

ICICI bank	101,511.87
------------	------------

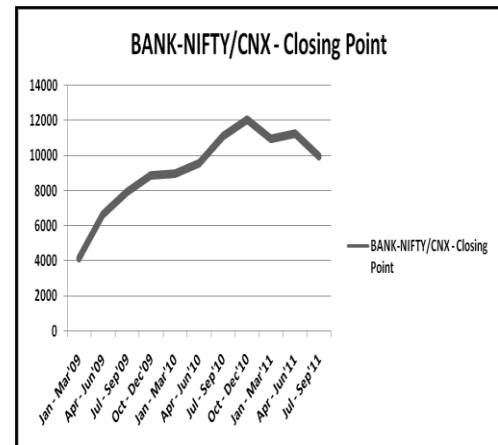
Figure 1. M Cap for Bank Nifty



From this above chart we came to know that those stocks are having high market capitalization. The top 5 stocks are

1. SBI 2. HDFC Bank 3. ICICI Bank 4. Axis Bank 5. Kotak Ban

Figure2. Nifty Month vs. closing point



Jul - Sep'09	7929.5	1806.623333	1464.743333	761.4083667	878.8992	688.9759333
Oct - Dec'09	8838.35	2259.822333	1709.802	879.5655	985.2665	786.4924667
Jan - Mar'10	8944.633333	2050.936667	1721.103667	870.9161333	1083.492667	775.5419333
Apr - Jun'10	9566.216667	2242.232667	1928.87	897.7605667	1224.306	755.3689
Jul - Sep'10	11091.23333	2729.337333	2170.513	981.1024333	1371.273	732.0604667
Oct - Dec'10	12024.93333	3045.063333	2331.660333	1152.891667	1435.747	480.7889
Jan - Mar'11	10927.55	2643.077667	2149.675333	1028.661	1282.679	411.9913
Apr - Jun'11	11249.75	2511.055333	2335.211333	1070.927	1294.388333	438.8183667
Jul - Sep'11	9965.116667	2178.83	770.122	951.5826667	1187.156	462.5889
Correlation	0.974366745	0.749795542	0.9907231	0.987016903	0.190982688	

Table 6. Beta Calculation for SBIN

Date	BANK Nifty /CNX Closing Point	SBIN Closing Price	SBIN Returns	BANK Nifty/CNX Returns	Beta SBIN
Jan - Mar'09	4160.716667	1088.290833			
Apr - Jun'09	6628.6	1500.755	37.90017838	59.31390024	0.7131334
Jul - Sep'09	7929.5	1806.623333	20.38096378	19.62556196	
Oct - Dec'09	8838.35	2259.822333	25.08541718	11.46163062	
Jan - Mar'10	8944.633333	2050.936667	-9.243455246	1.2025246	
Apr - Jun'10	9566.216667	2242.232667	9.327250474	6.949232136	
Jul - Sep'10	11091.23333	2729.337333	21.72409104	15.9416906	
Oct - Dec'10	12024.93333	3045.063333	11.56786287	8.418360449	
Jan - Mar'11	10927.55	2643.077667	-13.20122513	-9.125899495	
Apr - Jun'11	11249.75	2511.055333	-4.995022872	2.948510874	
Jul - Sep'11	9965.116667	2178.83	-13.23050626	-11.41921672	

Table 7. Beta Calculation for HDFC

Date	BANK Nifty/CNX Closing Point	HDFC Closing Price	HDFC Returns	BANK Nifty/CNX Returns	Beta HDFC
Jan - Mar'09	4160.716667	909.7766333			
Apr - Jun'09	6628.6	1278.691667	40.55006693	59.31390024	0.9422046
Jul - Sep'09	7929.5	1464.743333	14.55015863	19.62556196	
Oct - Dec'09	8838.35	1709.802	16.73048523	11.46163062	
Jan - Mar'10	8944.633333	1721.103667	0.660992735	1.2025246	
Apr - Jun'10	9566.216667	1928.87	12.07169196	6.949232136	

Jul - Sep'10	11091.23333	2170.513	12.52769756	15.9416906	
Oct - Dec'10	12024.93333	2331.660333	7.424389211	8.418360449	
Jan - Mar'11	10927.55	2149.675333	-7.804953295	-9.125899495	
Apr - Jun'11	11249.75	2335.211333	8.630884727	2.948510874	
Jul - Sep'11	9965.116667	770.122	-67.02131455	-11.41921672	

Table 8. Beta Calculation for ICICI

Date	BANK Nifty/CNX Closing Point	ICICI Closing Price	ICICI Returns	BANK Nifty/CNX Returns	Beta ICICI
Jan - Mar'09	4160.716667	377.7294667			
Apr - Jun'09	6628.6	590.2724667	56.26857811	59.31390024	0.883903615
Jul - Sep'09	7929.5	761.4083667	28.99269569	19.62556196	
Oct - Dec'09	8838.35	879.5655	15.51823417	11.46163062	
Jan - Mar'10	8944.633333	870.9161333	-0.983368118	1.2025246	
Apr - Jun'10	9566.216667	897.7605667	3.082321291	6.949232136	
Jul - Sep'10	11091.23333	981.1024333	9.283306673	15.9416906	
Oct - Dec'10	12024.93333	1152.891667	17.50981629	8.418360449	
Jan - Mar'11	10927.55	1028.661	-10.7755716	-9.125899495	
Apr - Jun'11	11249.75	1070.927	4.108836633	2.948510874	
Jul - Sep'11	9965.116667	951.5826667	-11.14402133	-11.41921672	

Table 9. Beta Calculation for AXIS

Date	BANK Nifty/CNX Closing Point	AXISBANK Closing Price	AXIS-BANK Returns	BANK Nifty/CNX Returns	Beta AXIS-BANK
Jan - Mar'09	4160.716667	401.5554			
Apr - Jun'09	6628.6	645.7235667	60.8055991	59.31390024	0.921253039
Jul - Sep'09	7929.5	878.8992	36.11075162	19.62556196	
Oct - Dec'09	8838.35	985.2665	12.1023321	11.46163062	
Jan - Mar'10	8944.633333	1083.492667	9.969502363	1.2025246	
Apr - Jun'10	9566.216667	1224.306	12.99624236	6.949232136	
Jul - Sep'10	11091.23333	1371.273	12.00410682	15.9416906	
Oct - Dec'10	12024.93333	1435.747	4.701762523	8.418360449	
Jan - Mar'11	10927.55	1282.679	-10.66120981	-9.12589949	
Apr - Jun'11	11249.75	1294.388333	0.912881009	2.948510874	
Jul - Sep'11	9965.116667	1187.156	-8.284402004	-11.4192167	

Table 10. Beta Calculation for KOTAK

Date	BANK Nifty/CNX Closing Point	KOTAK Closing Price	KOTAK Returns	BANK Nifty/CNX Returns	Beta KOTAK-BANK
Jan - Mar'09	4160.716667	401.5554			
Apr - Jun'09	6628.6	645.7235667	60.8055991	59.31390024	0.921253039
Jul - Sep'09	7929.5	878.8992	36.11075162	19.62556196	
Oct - Dec'09	8838.35	985.2665	12.1023321	11.46163062	
Jan - Mar'10	8944.633333	1083.492667	9.969502363	1.2025246	
Apr - Jun'10	9566.216667	1224.306	12.99624236	6.949232136	
Jul - Sep'10	11091.23333	1371.273	12.00410682	15.9416906	
Oct - Dec'10	12024.93333	1435.747	4.701762523	8.418360449	
Jan - Mar'11	10927.55	1282.679	-10.66120981	-9.125899495	
Apr - Jun'11	11249.75	1294.388333	0.912881009	2.948510874	
Jul - Sep'11	9965.116667	1187.156	-8.284402004	-11.41921672	

Figure 3. Correlation - Nifty Vs Top Banks

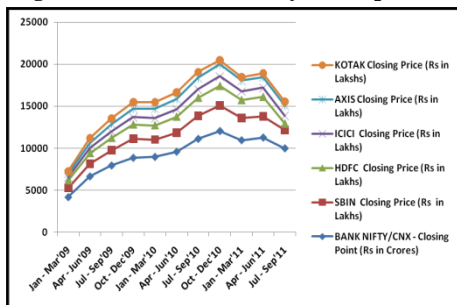


Figure 4. SBI - Beta Calculation

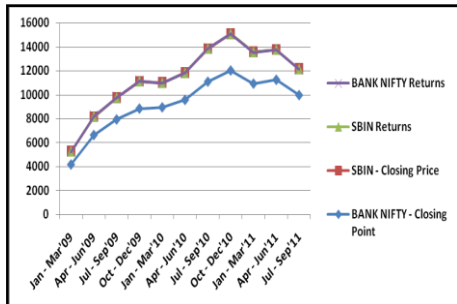


Figure 5. HDFC - Beta Calculation

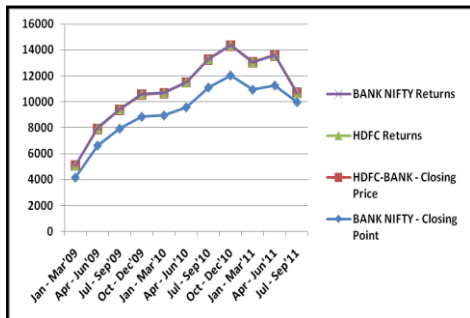


Figure 6. ICICI - Beta Calculation

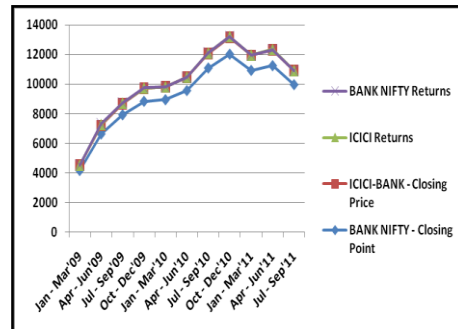


Figure 7. AXIS - Beta Calculation

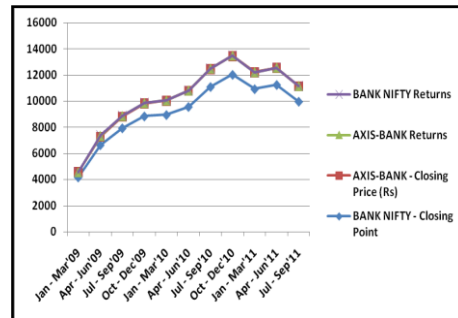
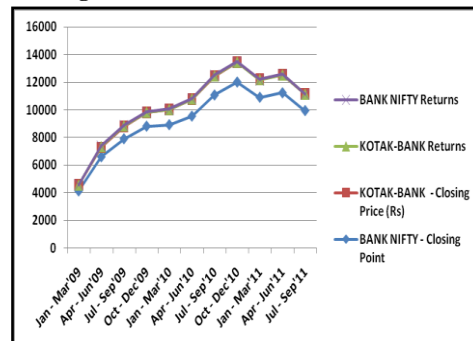


Figure 8. KOTAK - Beta Calculation



6. CONCLUSION

Many Clear studies were made based on the historical data of the NSE stock market to identify the flow and changes which happens in stock market for each index. From the three years data, which had commence from January 2009 and up to the first week of September 2011. The graph which gives a notion about the various changes happened in the stock market is shown. The above graph shows a sudden fall in the stock market during the recession period. The flow shows a constant growth after 2009 which gives positive sign for regular investors and also for naïve investor. We wish to conclude from the paper that investing in the banking index in stock market will always give profitable solutions to the naives investors. It is an innovative beginning and proposed system to introduce the decision making power for the naive investors through the next research work.

7. REFERENCES

- [1]. Basaltoa N, Bellottib R, De Carlob F, Facchib P, Pascazio S (2005). Clustering stock market companies via chaotic map synchronization, *Physica A*.
- [2] Chi-Lin L, Ta-Cheng C (2009). A study of applying data mining approach to the information disclosure for Taiwan's stock market investors, *Expert Systems with Applications*.
- [3] David E, Suraphan T (2005). The use of data mining and neural networks for forecasting stock market returns, *Expert Systems with Applications*.
- [4] Hsiao-Fan W, Ching-Yi K (2004). Factor Analysis in Data Mining, *Computers and Mathematics with Applications*.
- [5] Jar-Long W, Shu-Hui C (2006). Stock market trading rule discovery using two-layer bias decision tree, *Expert Systems with Applications*.
- [6] Jie S, Hui L (2008). Data mining method for listed companies' financial distress prediction, *Knowledge-Based Systems*.
- [7] R. J. Povinelli, "Time Series Data Mining: Identifying Temporal Patterns for Characterization and Prediction of Time Series Events", Ph.D. Dissertation, Marquette University, 180 p., 1999.
- [8] X. Feng and H. Huang, "A Fuzzy-Set-Based Reconstructed Phase Space Method for Identification of Temporal Patterns in Complex Time Series", *IEEE Trans. on Knowledge and Data Engineering*, Vol. 17, No. 5, pp. 601-613, 2005.
- [9] R. J. Povinelli and X. Feng, "Data Mining of Multiple Nonstationary Time Series, in *Proceedings of Artificial Neural Networks in Engineering*", St. Louis, Missouri, pp. 511-516, 2005.
- [10] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge:Cambridge University Press, 388 p., 1997.
- [11] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, SanFrancisco: Academic Press, 800 p.,2007
- [12] S.M. Weiss and N. Indurkha, *Predictive Data Mining: A practical Guide*. San Fransisco: Morgan Kaufmann., 228 p., 1998.
- [13]. clustering stock market companies via chaotic map synchronization –Elsevier –*physica A* 345 –(2005)-196-206.
- [14].A heuristic forecasting model for stock decision making – *Math ware and soft computing* – 2005 – 33-39.
- [15].A review on Data mining Applications to the performance of Stock marketing – *IJCA-(0975-8887)* volume 1 – No.3
- [16].world Federation of Exchanges "2007 "Annual report and market statistics – Retrieved on 2008-09-30
- [17].Fama E.F., (1970) Efficient capital markets : Areview of theory and empirical work " a journal of finance 25.pp-384-417
- [18]. Han:,Jiawei and kamber Micheline 2006 : Mining concepts and techniques –second edition
- [19].Eng.W.F.1988.the technical analysis of stocks options and futures –advanced trading systems and techniques
- [20].An approach for generating financial forecasting using data mining techniques – *ICETT-2011*.
- [21].Analysis and clustering of nifty companies of share market using data mining tools –*journal of Engineering Research and studies*.
- [22].Technical market indicators optimization using evolutionary Algorithms – *GECCO 2008 - ACM - 978-1-60558-131*[19]"News Headlines" ([http:// www. cnbc. com/ id/ 27166818](http://www.cnbc.com/id/27166818)). Cnbc.com. October 13, 2008. Retrieved March 5, 2010
- [23] Learn About Indian Stock Market - & How Shares are Traded in India for NRIs,OCIs and PIOs?? By *NriInvestIndia.com Dated: May 24, 2008*