Data Characterization for Study and Analyze Performance of Different Sectors of Indian Market

Prof. Pareshkumar J. Patel
U.V.Patel College of Engineering,
Ganpat University, Kherva
Gujarat, India

Dr. Narendra J. Patel
U.V.Patel College of Engineering
Ganpat University, Kherva
Gujarat, India

Abstract — **It has become essential to find better and effective way for analysis of growth in investment. There is a vast scope for investment in diverse instruments in Indian Stock Market. Before investment it is important to identify characteristics of indices using its past performance. The purpose of this paper is to present trend of specific index and scope of investment in selected sector using various Data Mining techniques. The technical analysis gives basic idea about the past and future trend and profitability of the investment.**

Keywords: Data Mining, Characterization, Concept Hierarchy, Class Comparison, Financial analysis

I. INTRODUCTION

In this paper we emphasis on various dimensions and their characterization using various comparative technical methods. Two different indices of National Stock Exchange CNX-Nifty and CNX-IT are studied for transactions on them during 1st January 2005 to 31st December 2011. The data is collected from official site www.nseindia.com. It is a secondary data which is reliable because it is provided by the Government authorized stock exchange site. The data is collected and organized day wise. Each and every transaction is classified as Monthly average value and summarized upto year wise group using mean of the value. Concept Hierarchy is applied on selected dimensions in both the classes of data. This data is characterized and interesting dimensions with applied concept hierarchy is selected for further analysis. The summarized data of CNX-Nifty class is compared with CNX-IT. The technical analysis using various methods proves that there is a significant growth in the volume of shares traded on CNX-Nifty index over CNX IT.

II. INDICES ANALYZED

A. S&P CNX Nifty Index

The CNX Nifty is India's stock exchange of selected 50 stocks from diversified sectors. It is used for a variety of purposes such as benchmarking fund portfolios, index based derivatives and index funds. The Exchange has also introduced trading in Futures and Options contracts based on CNX-IT, BANK NIFTY, and NIFTY MIDCAP 50 indices [1] [2].

The S&P CNX Nifty Index represents about 66.90% of the free float market capitalization of the stocks listed on NSE as on December 30, 2011. The total traded value for the last six months of all index constituents is approximately 56.58% of the traded value of all stocks on the NSE. Impact cost of the S&P CNX Nifty for a portfolio size of Rs.50 lacks is 0.08%.

B. CNX IT Index

Information Technology (IT) industry has played a major role in the Indian economy during last few years. A number of large, profitable Indian companies today belong to the IT sector and a great deal of investment interest is now focused on the IT sector [2].

Companies in this index are those that have more than 50% of their turnover from IT related activities like IT Infrastructure, IT Education and Software Training, Telecommunication Services and Networking Infrastructure, Software Development, Hardware Manufacturer's, Vending, Support and Maintenance [1].

III. DATA MINING TECHNIQUES

Data mining techniques are used to discover hidden knowledge, unknown patterns and new rules from large data sets, which maybe useful for a variety of decision making activity. With the increasing economic globalization and improvements in information technology, large amounts of financial data are being generated and stored. These can be subjected to data mining techniques to discover hidden patterns and obtain predictions for trends in the future and the behaviour of the financial markets [3]. With the immediacy offered by data mining, latest data can be mined to obtain crucial information at the earliest. This in turn would result in an improved market place responsiveness and awareness leading to reduced costs and increased revenue.

There are many techniques of Data Mining for identifying useful patterns and analysis. The main two categories of Data Mining techniques are Predictive Data Mining and Descriptive Data Mining. Descriptive Data Mining describes data in concise and summarative manner and presents interesting general properties of data. Summarization can be presented using Class Description. There are main two ways of Data

Summarization one is Characterization and the other is Class Comparison.

A. *Characterization*

Data characterization is a summarization of general features of values in a target class, and produces what is called *characteristic rules*. The data relevant to a user-specified class are normally retrieved by a database query and run through a summarization module to extract the essence of the data at different levels of abstractions [4]. With concept hierarchies on the attributes describing the target class, the *attribute-oriented induction* method can be used, for example, to carry out data summarization.

B. *Class Comparison*

Class Comparison is used for Data discrimination. It produces what are called *discriminant rules* and is basically the comparison of the general features of objects between two classes referred to as the *target class* and the *contrasting class* [4]. The techniques used for data discrimination are very similar to the techniques used for data characterization with the exception that data discrimination results include comparative measures.

IV. DATA PREPROCESSING

Data is collected from the official website of National Stock Exchange of India between 01/01/2005 and 31/12/2011 for all the working days. This data is downloaded in CSV format. Selected data from two different sectors CNX Nifty and CNX IT are organized in a database as per their transaction on specific date. It was auto generated data from all the transactions and there was no outlier and missing data in it. Data focusing is done for selection of filtered data from the database. The generated initial working set is processed for attribute removal as per the rules of Attribute Oriented Induction. Irrelevant attributes for further analysis like High, Low, Open, etc. are removed from the data and only the attributes required to analyze are kept. The selected attributes kept for further processing are Date of transaction, CNX-Nifty Value, CNX-Nifty Turnover, CNX-IT Value and CNX-IT Turnover.

Here the data is in huge quantity for 6 years. So, it is reduced by taking monthly average of it. Concept Hierarchy is generated on the date of transaction and the values are aggregated monthly and yearly. Then attribute interestingness is checked, based upon the interestingness of the attributes generalization is performed and the Prime Generalize Relation is obtained using attribute oriented induction. Now the data was processed and ready to use is shown in Table-1 Prime Generalized Relation. It is used for further analysis of Turnover and Value of the selected indices.

TABLE I. PRIME GENERALIZED RELATION

Year	CNX-Nifty-Value	CNX-IT Value	CNX-Nifty Turnover (Rs. Cr)	CNX-IT Turnover (Rs. Cr)
2005	2268.91	3066.22	2747.20	672.03
2006	3357.09	4337.71	3989.28	794.26
2007	4571.29	5054.30	5397.45	869.30
2008	4339.11	3633.18	6973.07	794.60
2009	4113.96	3738.89	8582.37	1090.97
2010	5461.12	6203.26	6606.59	818.56
2011	5335.91	6375.93	5998.54	841.31

V. APPLYING CHARACTERIZATION

Characterization rule is applied on the prime generalized relation to identify difference in the turnover of CNX-Nifty and CNX-IT.

TABLE II. TURNOVER COMPARISON

Year	2005	2006	2007	2008	2009	2010	2011
CNX-Nifty Turnover (Rs. Cr)	2747.20	3989.28	5397.45	6973.07	8582.37	6606.59	5998.54
CNX-IT Turnover (Rs. Cr)	672.03	794.26	869.30	794.60	1090.97	818.56	841.31
Total Turnover (Rs. Cr)	3419.23	4783.54	6266.75	7767.67	9673.34	7425.15	6839.85

A quantitative characteristic rule can be represented by associating the corresponding t-weight value with each disjunct covering the target class. The value of t-weight is [0.0, 1.0] or [0%, 100%].

$$\forall X, t\,arg\,et\_class(X) \Rightarrow condition_1(X)[t:w_1]$$
$$\vee ... \vee Condition_m(X)[t:w_m]$$

(1)

Applying the quantitative characteristic rule to the data given in Table-II, we get following rules. The t-weight value of CNX-Nifty and CNX-IT indicates turnover of specific index value. Here the rule is applied for the value of year 2005 is shown in (2) and 2011 is shown in (3) to identify difference between the weightage of turnover in selected years values. Here CNX-Nifty and CNX-IT is taken as conditions for specific year.

$$\forall X, Year(X) = "2005" \Rightarrow$$
$$(Index(X) = "CNX - Nifty")[t:80.34\%]$$
$$\vee (Index(X) = "CNX - IT")[t:19.66\%]$$

(2)

$$\forall X, Year(X) = "2011" \Rightarrow$$
$$(Index(X) = "CNX - Nifty")[t:87.69\%]$$
$$\vee (Index(X) = "CNX - IT")[t:12.31\%]$$

(3)

Above characteristic rules indicates weightage of CNX-Nifty and CNX-IT in total turnover of both the indices. It indicates the difference is increased from the year 2005 to 2011. It indicates variations in the turnover and also that turnover of CNX-Nifty is increasing over time. There is no such significance difference in turnover of CNX-IT during this time period.

Turnover Analysis is shown in Figure 1. Turnover of CNX-Nifty clearly indicates the rise in the turnover and CNX-IT's turnover is increased but very low in comparison to CNX-Nifty. It shows increase in trading volume of CNX-Nifty.

Value comparison of CNX-Nifty and CNX-IT is shown in Figure 2. It shows that the value of both the indices have no major difference in comparison to their turnover.
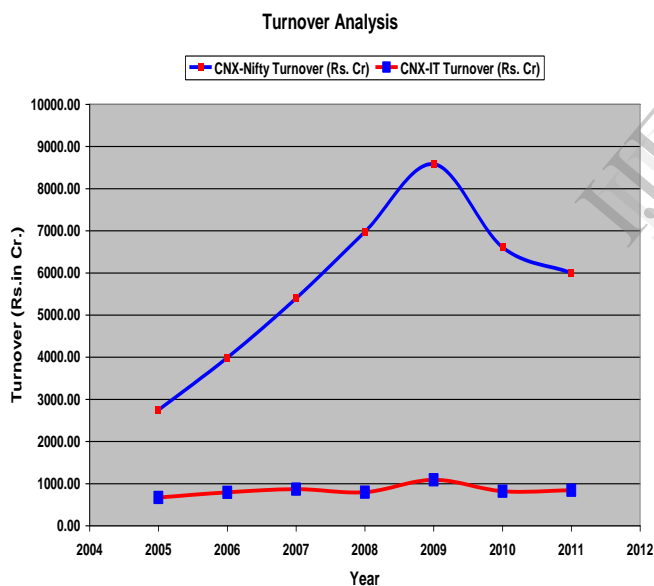

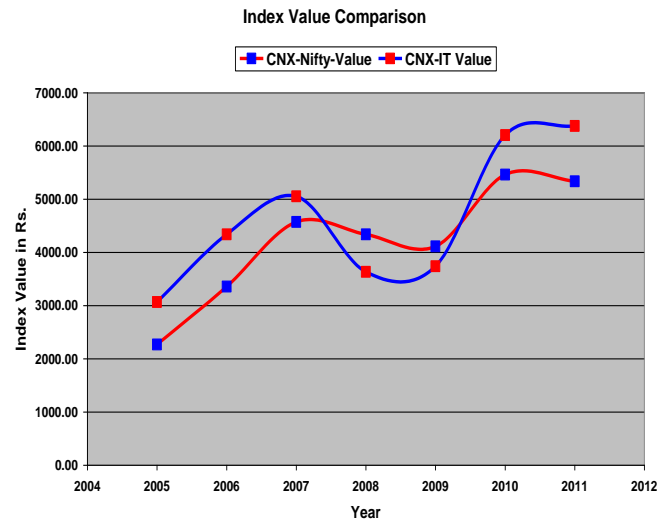Figure 1. Turrnover Analysis of CNX-Nifty and CNX-IT

Figure 2. Value comparison of CNX-Nifty and CNX-IT

VI. RESULT ANALYSIS

Analysis of indices CNX-Nifty and CNX-IT by value is shown in Figure 2 and it indicates that there is no significant difference between the values of the indices in comparison to their differences in turnover during the time duration of six years. If the investor would have invested in any of the indices he could have got same benefit from both the indices.

Turnover in both the indices shows higher difference between them. Analysis of sectors CNX-Nifty and CNX-IT by volume in (Cr.) is shown in the Table II and Figure I. It indicates that there is a vast difference between the turnovers of the indices during the time duration of six years. It shows that the CNX-Nifty has gained higher popularity over CNX-IT. Investors and traders are increasing on CNX-Nifty index in comparison to CNX-IT. Moreover there is a direct effect of recession in other countries and currency fluctuation on CNX-IT because it's main source of income is from foreign countries and profit is based on outsourcing. CNX-Nifty is a collection of stocks from different sectors. So, it does not have direct effect of recession in any one sector.

VII. CONCLUSION

Using data mining techniques to study financial transaction over long period of time is quite significant. This paper mainly studies the specific technique Characterization of data mining in the analysis of CNX-Nifty and CNX-IT indices listed in National Stock Exchange of India. Further more, it also analyses turnover and values of the indices using quantitative rules and finds weightage of each index.

The study of the indices CNX-Nifty and CNX-IT shows that the CNX-Nifty is a collection of various kind of stocks and is more popular among the traders of the shares in comparison to CNX-IT. Important point is the values of indices CNX-Nifty and CNX-IT have no major difference in comparison to their turnover. CNX-Nifty is a collection of stocks from various sectors whereas CNX-IT is a collection of IT sector stocks only. This analysis can be applied to analyze other sectors of Indian stock market like AUTO, BANK, FMCG, etc. for better investment perspective.

REFERENCES

[1] Vogas, New Delhi, India " Commodity Commodities Derivatives Technical Analysis Software Stock Forex Futures MCX SX NCDEX NMCE NSE BSE NSEL PXIL ICEX India", para. 2. [Online]. Available: http://www.vogaz.com/NseFo. [Accessed: Feb. 25, 2014].

[2] Government of India, "National Stock Exchange of India, Products Sectoral Indices CNX IT", para. 2. [Online]. Available: http://www.nseindia.com/products/content/equities/indices/sectoral_indices.htm. [Accessed: Aug. 25, 2013].

[3] Gabriella accomodation, "Analytics and Data Mining, Business Services", [Online]. Available:http://goranxview.blogspot.in/ [Accessed: Feb. 10, 2014].

[4] Jiawei Han, Shojiro Nishio, Hiroyuki Kawano, Wei Wang, "Generalization-based data mining in object-oriented databases using an object cube model", Data and Knowledge Engineering, Volume 25, Issues 1–2, Pages 55-97 (March 1998)

[5] Dou Xuexia, "Application of Data Mining Algorithms in the Analysis of Financial Distress Early Warning Model of Listed Company", Henan Polytechnic Zhengzhou, China xuexiad@163.com 978-1-61284-840-2/11/$26.00 ©2011 IEEE

[6] Rajanish Dass, "Data Mining In Banking And Finance: A Note For Bankers", Indian Institute of Management Ahmedabad

[7] J. Han and M. Kamber. "Data Mining: Concepts and Techniques". Morgan Kaufmann, 2000

[8] J. M. Zytkow and W. Klösgen, Rajnish Dass, "Data Mining in Banking and Finance : A note for bankers" Handbook of Data Mining and Knowledge Discovery. New York : Oxford