An Agglomerative Analysis of Nifty Companies for an Investment Perspective

Preeti Baser[1]

[1]Research Scholar, R. K. University, Faculty of
Science, Rajkot
Assistant Professor, MCA Programme,
Shri Jairambhai Patel Institute of Business
Management and Computer Applications,
Gandhinagar, Gujarat, India-382007;
email: preeti.dalal@gmail.com

Dr. Jatinderkumar R. Saini[2]

[2]Research Guide, R. K. University, Faculty of
Science, Rajkot.
Director (I/C) & Associate Professor,
Narmada College of Computer Application,
Bharuch, Gujarat, India;
email:saini_expert@yahoo.com

Abstract—In today's competitive financial market, individuals want to earn profit from their investments. Stock market is a public market where an individual can buy/sell a share at some level of risk and earn maximum profit from the investments. Data Mining is one of emerging field in computer science. It is a process of extracting knowledge and trends from large databases. Data mining techniques, like classification, clustering,association rule mining, and outlier analysis are widely used in various domain. Clustering is one of the data mining techniques is widely used in various financial domains. The objective of this research paper is to construct clusters of Nifty companies for better investment. This paper shows analysis of Nifty companies using hierarchical clustering algorithm which generates clusters based on some investment criteria. In this paper, price per earnings ratio and earnings per share is considered as an investment criteria. These clusters guide the investors to invest their funds in different securities.

Keywords:Clustering, Hierarchical Clustering, Price per earnings ratio,Earnings per share (EPS), Agglomerative Approach, Self Organizing Map(SOM)

_____ ***** _____

I. INTRODUCTION

In today's competitive financial market, an individual wants to earn profit from his/her investments. Stock market is a public market where an individual can buy/sell a share at some level of risk and earn maximum profit from the investments. There are various data mining techniques available, like classification, clustering,association rule and outlier analysis. Data mining techniques are also widely applied in various areas of stock market, including predicting stock prices, predicting stock indices, portfolio management, portfolio risk management, trend detection, designing recommender systems, etc.[1,2]. Clustering is one of most popular data mining technique is widely used in various financial domains. It is an unsupervised technique in which class label is not available. It is a group of objects that share the common characteristics. The main objective of clustering is that the data objects within a group be similar to one another and different from the objects in the other groups. The greater the similarity within a group and greater the difference between groups, the better or more distinct is the clustering. This paper shows analysis of Nifty companies using hierarchical clustering algorithm. Hierarchical clustering techniques that generate clusters based on some investment criteria. In this paper, price per earnings (P/E) and EPS ratios have been consideredas investment criteria. These clusters guide investors to invest their money in securities. P/E ratio and EPS data are collected for 50 Nifty companies for the current year.

A. STOCK MARKET

The stock market is an electronic trading platform for the investors to trade the equity shares of the listed companies. It gives ample opportunity to the investors for optimal and profitable investment along with full liquidity. While trading, an investor looks at various financial and other aspects of the companies whose shares he/she is planning to buy or sell. In stock market, share prices are dependent on so many factors that affect the price of stocks [3]. Among these, the P/E and EPS ratios are very popular and widely used by the investors as an investment criterion.

B. PRICE PER EARNINGS RATIO

The P/E ratio gives a fair idea of how a company's share price compares with its earnings. If the price of the share is too much lower than the earnings of the company, the stock is undervalued and it has the potential to rise in the near future. On the other hand, if the price is way too much higher than the actual earnings of the company, then the stock is said to overvalued and the price can fall at any point. The most commonly used guide to the relationship between stock prices and earnings is the P/E ratio. The P/E ratio is volatile and may fluctuate considerably. The P/E ratios (above 20, thumb rule) are characteristic of growth companies, although with the average market multiple currently around 28, a P/E ratio of 20 almost seems like a value stock. High P/E ratios indicate high risk. If the future anticipated growth of the high P/E ratio stocks is not achieved, their stock prices will be punished and they will fall very quickly. On the other hand, if they live up to their promise, investors will benefit substantially. Low P/E ratio stocks (under 10) are characteristic of either mature company with low growth potential, or companies that are undervalued or in financial difficulty. By comparing the P/E ratio of a company with the averages in the industry and the markets, investors can get a feeling for the relative value of the stock. P/E ratios fluctuate considerably, differing among companies due to many factors, from growth rates and popularity to earnings and other financial characteristics. It

is calculated using the equation: P/E ratio = market price of the stock/earnings per share.

C. EARNINGS PER SHARE (EPS):

It refers to the ratio of the profit after tax of the company for any financial year after payment of preference dividend.The equity shareholders are the sole claimants to the net earnings of the corporation after making payment of dividend to the preference shareholders. The significance of this ratio flows from the fact that higher the earnings per share the more is the scope for a higher rate of dividend and also of retained earnings, to build up the inner strength of the company.

$$EPS = \frac{(Net\ Profit\ After\ Tax - Preference\ Dividend\ )}{No.of\ outstanding\ shares}$$

D. CLUSTER ANALYSIS

Cluster Analysis, also called data segmentation, has a variety of goals. All relate to grouping or segmenting a collection of objects (also called observations, individuals, cases, or data rows) into subsets or "clusters", such that those within each cluster are more closely related to one another than objects assigned to different clusters. Central to all of the goals of cluster analysis is the notion of degree of similarity (or dissimilarity) between the individual objects being clustered. It is also referred to as an unsupervised technique. An unsupervised technique means that the class labels are not available. In this research paper, hierarchical clustering technique is carried out to cluster stock market data. Clustering of stock market data can be carried out based on some criteria.

II. LITERATURE SURVEY

In [4], authors discussed about the application of cluster analysis in biological field. They analyzed K-means and Fuzzy C-Means algorithms for identification of hematoma in brain CT scan. The behavior of both the algorithms depends upon the brain CT scan images. They investigated that K-means algorithm gives clean clusters as compared to Fuzzy C-Means. The authors [5] examined the use of cluster analysis in accounting domain to detect discrepancy in audit. In [6], authors reviewed applications of cluster analysis in agriculture field. They discussed about how cluster analysis tools are useful for improving agricultural management, predicting and suggesting solutions in agricultural areas. In [7] represents application of K-Means clustering algorithm to predict student's academic performance. The authors in [8] used SOM to define best companies or group of best companies of USA stock markets data for investment. In this work, fundamental financial parameters were used as inputs, and the output of SOM formed the different group of companies located into clusters. The cluster analysis is also widely used in stock market prediction [9]. The authors in [10] have discussed various clustering techniques and its applications in various domains. The purpose of paper [11] is to explain detailed analysis of agglomerative hierarchical clustering algorithm on Bulgarian Stock Exchange data to identify similar temporal behavior of the traded stocks.

III. METHODOLOGY

In hierarchical clustering the data are not partitioned into a particular cluster in a single step. Instead, a series of partitions takes place, which may run from a single cluster containing all objects to n clusters each containing a single object. Hierarchical Clustering is subdivided into agglomerative methods, which proceed by series of fusions of the n objects into groups, and divisive methods, which separate n objects successively into finer groupings. Agglomerative techniques are more commonly used. Hierarchical clustering may be represented by a two dimensional diagram known as dendrogram.

- AGGLOMERATIVE METHODS
  Steps:
1. Place each object in its own cluster.
2. Find similarity or dissimilarity of a pair of objects in the dataset.
3. Merge pair of cluster into larger and larger clusters by using four criteria (single linkage, complete linkage, average linkage).
4. Continue steps 2 and 3 until all objects are in a single cluster.

Similarity between pairs of objects can be calculated fromsingle linkage, complete linkage, and average linkagemetrices [12].

a) Single linkage clustering

Single linkage is also known as the nearest neighbor technique. The distance between groups is defined as the distance between the closest pair of objects, where only pairs consisting of one object from each group are considered and merged.

b) Complete linkage clustering

The complete linkage, also called farthest neighbor, clustering method is the opposite of single linkage. Distance between groups is now defined as the distance between the most distant pair

c) Average linkage clustering

Here the distance between two clusters is defined as the average of distances between all pairs of objects, where each pair is made up of one object from each group.

d) Average group linkage

With this method, groups once formed are represented by their mean values for each variable, that is, their mean vector, and inter-group distance is now defined in terms of distance between two such mean vectors.

IV. DATA DESCRIPTION AND INTERPRETATION

Data for the financial year 2013 are collected from the financial databases of 50 Nifty companies. Annexure 1 shows 50 Nifty company names, P/E and EPS ratios. The Hierarchical algorithm is applied to cluster Nifty companies containing some financial information. The numbers of clusters are 6 and following clusters are generated.

TABLE 1 MINIMUM DISTANCES BETWEEN EACH PAIR OF CLUSTER OBJECTS

Stage	Cluster 1	Cluster 2	Distance
1	18	32	0.280664
2	35	36	0.967125
3	20	38	1.298655
4	11	25	1.557391
5	12	29	1.609627
6	11	12	1.747166
7	17	31	2.527351
8	20	24	2.615633
9	5	19	2.681281
10	9	11	2.755634
11	9	50	2.783617
12	10	15	3.019522
13	8	10	3.509769
14	17	33	4.088834
15	2	47	4.382649
16	20	35	4.75857
17	18	22	5.11657
18	4	5	5.387893
19	2	9	5.451236
20	7	8	5.858737
21	2	18	6.124623
22	2	21	6.132135
23	2	37	6.499859
24	2	7	6.701013
25	4	6	6.905536
26	2	28	6.990353
27	23	34	7.234357
28	23	41	7.583131
29	2	27	8.135952
30	2	20	8.530355
31	4	49	8.620092
32	1	45	8.792183
33	1	17	8.838839
34	1	23	8.872751
35	1	14	8.874187
36	1	48	8.995309
37	2	30	9.214278
38	2	3	11.691947
39	16	39	15.727213
40	1	2	15.957358
41	1	40	16.36862
42	1	4	17.916798
43	1	16	24.488927
44	1	26	25.39652
45	13	42	32.577516
46	1	13	35.491982
47	1	43	48.083414
48	1	44	52.071712
49	1	46	120.579935

This output gives information about the cluster formation. Initially, each individual case is considered its own cluster (with just itself as a member), so it starts with number of clusters = number of data. At stage 1, above, clusters 18 and 32 have been found to be closer together than any other two clusters, so they are joined together in a cluster called Cluster 18. So now we have one cluster that has two cases (cases 18 and 32), and 48 other clusters that still have just one case in each. At stage 2, clusters 35 and 36 are found to be closer together than any other two clusters, so they are joined together into cluster 35. The cluster ID is thus the lowest case number of the cases belonging to that cluster. This process continues until there is just one cluster. At various stages of the clustering process, there are different numbers of clusters. Graphs called a dendrogram visualize.
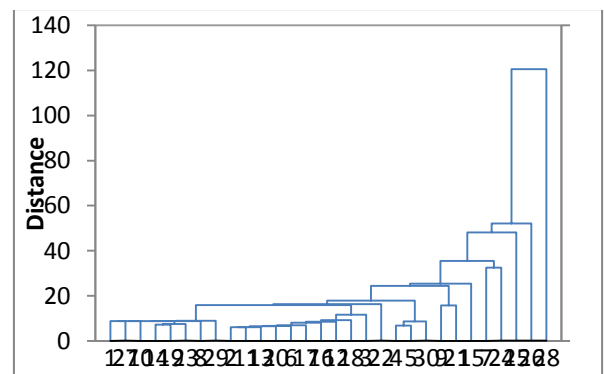

FIGURE-1 DENDROGRAM OF CLUSTERING RESULTS

In the above dendrogram, the Sub Cluster IDs are listed along the x-axis. The y-axis measures inter-cluster distance. Consider cases 4 and 5 have an inter-cluster distance of 5.38. No other cases have a smaller inter-cluster distance, so 4 and 5 are joined in a cluster, indicated by the horizontal line linking them. The horizontal line through the diagram at any level on the y-axis, the vertical cluster lines it intersects indicates clusters whose members are at least that close to each other.

V. RESULT AND DISCUSSION

TABLE 2 CLUSTER INFORMATION OF NIFTY COMPANIES

Cluster Id	Companies
Cluster-1	ACC Ltd., Hindustan Unilever Ltd.,Ambuja Cements Ltd.,Housing Development Finance Corporation Ltd.,Asian Paints Ltd.,ICICI Bank Ltd.,Axis Bank Ltd., IDFC Ltd., Bajaj Auto Ltd., IndusInd Bank Ltd., Bank Of Baroda,InfosysLtd.,Bharat Heavy Electricals Ltd.,ITCLtd.,Bharat Petroleum Corporation Ltd.,Jaiprakash Associates Ltd.,BhartiAirtelLtd.,Jindal Steel & Power Ltd.,Cairn India Ltd.,Kotak Mahindra Bank Ltd.,CiplaLtd.,Larsen& Toubro Ltd.,Coal India Ltd.,Lupin Ltd., DrReddys Laboratories Ltd.,Mahindra& Mahindra Ltd.,GAIL (India) Ltd.,Maruti Suzuki India Ltd.,Grasim Industries Ltd.,NMDCLtd.,HCL Technologies Ltd.,NTPCLtd.,HDFC Bank Ltd.,Oil& Natural Gas Corporation Ltd.,HeroMotoCorpLtd.,Power Grid Corporation Of India Ltd.,Hindalco Industries Ltd.,Punjab

1350

	National Bank,Tata Consultancy Services Ltd.,Ranbaxy Laboratories Ltd.,Tata Power Company Ltd.,Reliance Industries Ltd.,Tata Steel Ltd.,Ultratech Cement Ltd.,Wipro Ltd.,
Cluster-2	DLF Ltd.
Cluster-3	SesaSterlite Ltd.
Cluster-4	State Bank Of India
Cluster-5	Sun Pharmaceutical Industries Ltd.
Cluster-6	Tata Motors Ltd.

VI. CONCLUSION

Finally, the information on the distance between the final clusters to evaluate the cluster validity. The ratio of the sum of squared distances for a given *k* to the sum of squared distances to the mean of the records is a useful measure for the usefulness of the clustering. If the ratio is near 1.0, then the clustering has not been very effective, whereas if it is small, we have well-separated groups. It is clear that cluster 1 has majority companies and this group has 55 companies; so, in future, we can construct a portfolio of this group instead of constructing portfolios of the remaining clusters, as this cluster is most accurate and compact. Clusters 2, 3, 4, 5 and 6 have only one company, so a diversified portfolio will not be generated. Hence, clusters 2, 3, 4, 5 and 6 are not considered as effective clusters.

REFERENCES

[1] S. Kaur, V.Mangat, 2012, "Applications of Data Mining in Stock Market", Journal of Information and Operations ManagementVol- 3, Issue-1, pp. 86-88.

[2] S. Kambey, R.S. Thakur &S.Jalori, 2012 "Application of Data Mining Technique in Stock Market: An Analysis", International Journal of Computer &Communication Technology (IJCCT), , Vol-3, Issue-3 pp.52-53.

[3] A. Malhotra, "Factors Affecting Stock Prices" retrieved from http://ezinearticles.com/?Factors-Affecting-Share-Prices&id=741674

[4] B.Sharma, K.Venugopalan, 2012, "Performance Evaluation of K-Means and Fuzzy C-Means Clustering Algorithms for Identification of Hematoma in Brain CT scan Images", International Journal of Advanced Research in Computer Science, Vol-3, Issue-2, pp. 219-222.

[5] S. Thiprungsri, MiklosA.Vasarhelyi, 2011,"Cluster Analysis for Anomaly Detection in Accounting Data : An Audit Approach", International Journal of Digital Accounting Reseach,Vol-11,Issue-11 pp. 69-84.

[6] M. Tiwari, B.Misra, 2011, "Application of Cluster Analysis in Agriculture- A Review Article", International Journal of Computer Applications, Vol-36, Issue-4, pp. 43-47.

[7] O.J. Oyelade et.al, 2010, "Application of K-Means Clustering Algorithm for Prediction of Students' Academic Performance", International Journal of Computer Science and Information Security(IJCSIS), Vol-7, Issue-1,pp. 292-295.

[8] G. Stankevicius, 2001. "Forming of the Investment Portfolio Using the Self-organizing Maps(SOM)", Informatica,Vol-12, Issue-4, pp. 573-584.

[9] M. Suresh, N. Geethanjali and B. Satyanarayan, 2012, "Clustering Approach to Stock Market Prediction", International Journal of Advanced Networking and Applications, Vol-3, Issue-4, pp.1281-1291.

[10] N. Mehta S. Dang"A Review of Clustering Techniques in various Applications for effective data mining" International Journal of Research in Engineering & Applied Science vol. 1, No. 1 2011.

[11] V. Marinova, 2008, "Usingthe Agglomerative Method of Hierarchical Clustering As a Data Mining Tool in Capital Market" International Journal of Information Theory and Applications, Vol-15.

[12] B. Rama et. al , 2010, "A Survey on clustering Current Status and challenging issues"(IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 9, pp. 2976-2980.

ANNEXTURE 1

TABLE-2 NIFTY COMPANIES WITH THEIR P/E AND EPS RATIOS

Sr.No.	Company Name	P/E Ratio	EPS(Rs.)
1	ACC Ltd.	25.29	56.52
2	Ambuja Cements Ltd.	23.89	8.41
3	Asian Paints Ltd.	44.89	10.95
4	Axis Bank Ltd.	11.76	110.68
5	Bajaj Auto Ltd	17.07	105.18
6	Bank Of Baroda	6.37	106.37
7	Bharat Heavy Electricals Ltd.	6.55	27.03
8	Bharat Petroleum Corporation Ltd.	10.35	36.55
9	BhartiAirtel Ltd.	21.74	13.42
10	Cairn India Ltd.	8.02	33.93
11	Cipla Ltd.	20.23	18.77
12	Coal India Ltd.	19.94	15.51
13	DLF Ltd.	79.47	2.95

14	DrReddys Laboratories Ltd.	23.7	74.53
15	GAIL (India) Ltd.	10.07	31.71
16	Grasim Industries Ltd.	21.05	133.58
17	HCL Technologies Ltd.	14.98	51.82
18	HDFC Bank Ltd.	22.08	28.27
19	Hero MotoCorp Ltd.	14.54	106.07
20	Hindalco Industries Ltd.	10.31	8.88
21	Hindustan Unilever Ltd.	26.55	17.56
22	Housing Development Finance Corporation Ltd.	26.34	31.35
23	ICICI Bank Ltd.	14.48	72.17
24	IDFC Ltd.	12.33	11.65
25	IndusInd Bank Ltd.	19.95	20.30
26	Infosys Ltd.	18.2	158.82
27	ITC Ltd.	32.92	9.39
28	Jaiprakash Associates Ltd.	29	2.26
29	Jindal Steel & Power Ltd.	20.44	17.04
30	Kotak Mahindra Bank Ltd.	35.74	18.23
31	Larsen & Toubro Ltd.	17.1	53.20
32	Lupin Ltd.	22.34	28.16
33	Mahindra & Mahindra Ltd.	15.16	56.80
34	Maruti Suzuki India Ltd.	16.16	79.21
35	NMDC Ltd	8.6	15.99
36	NTPC Ltd.	9.28	15.30
37	Oil & Natural Gas Corporation Ltd.	12.74	24.46
38	Power Grid Corporation Of India Ltd.	11.58	9.15
39	Punjab National Bank	5.34	134.31
40	Ranbaxy Laboratories Ltd.	0	-3.84
41	Reliance Industries Ltd.	11.89	65.04
42	SesaSterlite Ltd.	112.01	1.39
43	State Bank Of India	10.05	206.20
44	Sun Pharmaceutical Industries Ltd.	164.07	2.49
45	Tata Consultancy Services Ltd.	24.1	65.23
46	Tata Motors Ltd.	284.64	0.95
47	Tata Power Company Ltd.	22.32	4.32
48	Tata Steel Ltd.	5.99	52.13
49	Ultratech Cement Ltd.	19.29	96.85
50	Wipro Ltd.	19.06	22.94