

Agent based Stock Clustering for Efficient Portfolio Management

Preeti Baser

Research Scholar,

Faculty of Science, R.K. University, Rajkot 360020,
Gujarat, India

Jatinderkumar R. Saini, Ph.D.

Director (I/C) and Associate Professor,

Narmada College of Computer Application, Bharuch
392011, Gujarat India

ABSTRACT

This research paper proposed agent based framework for portfolio management using non-hierarchical clustering method. The framework included various agents such as data agent, clustering agent, ranking agent, portfolio manager and user agent. The data agent collected financial ratio of Nifty 50 companies from financial database. Clustering agents generated clusters and DB index computed to find optimum cluster size of each method. Validation agent evaluated the performance of k -means, k -medoids and fast k -means using intra-class inertia. Clusters generated by k -means used for investment and portfolio analysis using Markowitz model. This research helped to assemble a diversified portfolio of stocks with the use of clustering

Keywords

Clustering, Data mining (DM), Davis-Bouldin (DB) Index, Dunn Index, k -means, k -medoids, Partitioning Around Medoids(PAM), Silhouette index

1. INTRODUCTION

Data mining is a process of automatically discovering knowledge and predicting future trends from large financial markets. It creates opportunities for companies to make proactive and knowledge-driven decision in order to gain a competitive advantage. There are varieties of DM techniques available over past decades that include classification, similarity search, cluster analysis, association rule mining. Data mining techniques are also widely applied in number of financial areas, including predicting stock prices, predicting stock indices, portfolio management, portfolio risk management, trend detection, designing recommender [27, 28].

Portfolio management is one of major problem in financial domain. In today's competitive financial environment, an investor wants to earn maximum profit from his assets. An investor considers an investment in securities faces with the problem of choosing from among a large number of securities. He confuses in which security he has to invest. It depends upon the risk-return characteristics of individual securities. He selects most desirable securities and likes to allocate his funds over this group of securities. Again, he faces with the problem of deciding which securities to select and how much to invest in each. The investor chooses the optimal portfolio taking into consideration the risk and return characteristics of all possible portfolios.

The research work describes about an agent based framework for portfolio management using non-hierarchical clustering methods. The proposed framework consist of various agents such as data agent, clustering agent, ranking agent, user agent and portfolio manager. This framework assists investors in strategic planning and investment decision-making. This research work can help to assemble a diversified portfolio of

stocks with the help of clustering and also will help investor community in specific and in turn it helps the society and economy in general for better allocation of wealth.

In this research paper, k -means, k -medoids and fast k -means clustering technique are applied to cluster financial ratio of Nifty 50 companies. Then each clustering algorithm is evaluated using intra-class inertia. Validity indices are used to find optimum number of cluster for clustering algorithm and to measure 'goodness' of clustering result by comparing to other ones which are created by other clustering algorithms, or by the same algorithms but using different parameter values. Several validity indices are presented here. If a dataset contains well-separated clusters, the distances among the clusters are usually large and the diameters of the clusters expected to be small [14]. Therefore larger value means better cluster configuration. Dunn index is defined as follows:

$$D = \min_{i=1 \dots n_c} \left\{ \min_{j=i+1 \dots n_c} \left(\frac{d(c_i, c_j)}{\max_{k=1 \dots n_c} (diam(c_k))} \right) \right\} \quad (1)$$

where $d(c_i, c_j) = \min_{x \in c_i, y \in c_j} \{d(x, y)\}$ and
 $diam(c_i) = \max_{x, y \in c_i} \{d(x, y)\}$.

The Davis-Bouldin (DB) [5] index measures the average of similarity between each cluster and its most similar one. As the clusters have to be compact and separated, the lower DB index value means better cluster configuration. Consequently, the number of clusters minimizing the DB index is recognized as the optimum quantity of clusters. For negative quantities of the DB index, the absolute amount is considered and lower amount have better quantities. Then DB index is defined as:

$$DB = \frac{1}{n_c} \sum_{i=1}^{n_c} R_i \quad (2)$$

where $R_i = \max_{j=1 \dots n_c, i \neq j} (R_{ij}), i = 1 \dots n_c$

One of the cluster validity index is Silhouette validation index. This index is defined:

$$S(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}} \quad (3)$$

$a(i)$ implies the average unlikeliness of i -observation of all other observations in the same cluster and $b(i)$ represents the minimum of average unlikeliness of i -observation to all observations in other segments (in the closest segment). Unlikeliness generally is regarded as the complement of similarity, and its result consists of the number of attributes that two objects uniquely have compared with the total number of attributes among them. If value of Silhouette index becomes near to 1, it is interpreted that all the objects in the sample are clustered well. On contrary, if value of Silhouette index becomes near to 0, it is deduced that objects could be arranged to the other cluster that has the same

distance with the current cluster. If value of Silhouette index becomes near to -1, it implies that sample point has been misclassified. It is merely somewhere among the clusters.

k-means clustering is a non-hierarchical and an unsupervised approach to form good clusters. The procedure starts with a certain *k*, the number of clusters. The main idea is to arbitrarily choose *k* objects from the entire dataset, one for each cluster, as the initial cluster center. The next step is to take each object from the given dataset and assign it to the cluster to which it is most similar, based on the distance between the object and the cluster center. Then, recalculate new *k* cluster centers. After obtaining new *k* cluster centers, new binding has to be done between the same data object and the new *k* cluster centers. This process repeats until there are no more changes in the *k* cluster centers. Typically, a squared error function is defined as follows:

$$\text{Squared Error}(SE) = \sum_{j=1}^k \sum_{i=1}^n \|x_i^j - c_j\|^2 \quad (3)$$

Where, $\|x_i^{(j)} - c_j\|^2$ is the chosen distance between data points x_i^j and cluster center c_j

The *k*-medoids method is mostly similar to *k*-means technique. It is also known as medoids shift algorithm. Both of the *k*-means and *k*-medoids techniques break the dataset into groups and intend to minimize squared error which is defined as the distance among object in a cluster and center of that cluster. The centroid of a cluster is always the one of the objects in the cluster. This is the major difference between the *k*-means and *k*-medoids. In the *k*-means algorithm the centroid of a cluster is frequently an imaginary point, not part of the cluster itself. *k*-medoids technique chooses objects as a medoids (centers). *k*-medoids is considered a partitioning approach of clustering as well that clusters the data set of *n* observations into *k* segments with *k* known a priori. The most prevalent realization of *k*-medoids clustering technique is Partitioning Around Medoids (PAM) [7]. *k*-means and fast *k*-means algorithms give exactly same result. Fast *k*-means algorithm avoids unnecessary distance calculations by applying triangle inequality in two different ways, and by keeping track of lower and upper bounds for distances between points and centroid. In contrast to the standard implementation of *k*-means, this implementation is much faster in many cases, especially for data sets with many attributes and a high *k* value, but it also needs more additional memory. For $k \geq 20$, it is many times faster than best previously known accelerated *k*-means method [33].

2. LITERATURE SURVEY

Sharma and Singh [29] studied the effect of fundamental factors on the equity prices of manufacturing firms listed on Bombay Stock Exchange. The study found that the book value, earnings per share and price-earnings ratio as significant share price determinants in engineering industry; book value and size in cotton textile industry; price per earnings ratio and return on capital employed in chemical industry; dividend, earnings and price per earnings ratio in electrical industry; and book value per share, payout and price per earnings ratio in miscellaneous industry. Mehta and Turan [26] identified market capitalization, market price to book value ratio and price per earnings ratio as major factors influencing share prices. Srinivasan [25] examined the fundamental determinants of share price in India for panel data over the period 2006-2011 in different six sectors. The empirical results revealed the dividend per share, earnings

per share and price per earnings ratio being the crucial determinants of share prices of these sectors.

Nirmala et al. [24] identified the determinants of share prices in the Indian stock market. The study focused on three sectors, namely auto, health care and public sector undertaking over the period 2000-2009 and examined the effect of dividend, profitability, price per earnings ratio and leverage on share prices. Sharma [30] examined the empirical relationship between equity share prices of different industry groups and revealed that earnings per share, dividend per share and book value per share have a significant impact on the equity price of different industry groups in India. The literature by Malhotra [18] determined the factors that influence stock prices in the context of National Stock Exchange (NSE) 100 companies. The result indicated that firm's book value, earnings per share and price per earnings ratio had significant positive association with firm's stock price.

The literature [31] examined that there were two major issues related to evaluate cluster results using validity indices. First issue was how many clusters were present in available datasets and second issue was how much clustering algorithm was good. The main goal was to plot graph of indices with respect to cluster size and analyze this graph. Dunn index required lot of computation power and it was sensitive to noise also. Dunn index [9] combined diameters and dissimilarity between clusters to estimate most reliable cluster size. Silhouette index [13] used average distance to find dissimilarity between points to represent the structure of data and its cluster. Davis-Bouldin index [5] introduced concept of dispersion of cluster and dissimilarity between clusters to calculate index. Davis-Bouldin index was found to be best index among other indices according to [16]. The authors [20] compared various clustering techniques based on different criteria and suggested selection of appropriate clustering technique in different suitable domain. The authors [23] and [19] applied *k*-means and hierarchical clustering algorithm on Nifty stock data and grouped stock data based on financial ratios. In literature [22], authors explained various validity indices used to find optimum cluster size. They evaluated *k*-means algorithm using DB index and found optimum cluster size for *k*-means is three for Nifty stock data. In the study [21], authors proposed an agent based framework for portfolio management using clustering. This agent based framework comprised of various agents that helped to detect clusters automatically from stock data and assisted investors in strategic planning and investment decision-making.

3. RESEARCH METHODOLOGY

As shown in Figure 1, financial information of Nifty 50 companies is collected from the ACCORD FINTECH database for the year 2012-2013. These data are preprocessed before detailed analysis. Three clustering algorithms named *k*-means, *k*-medoids and fast *k*-means methods are applied on this dataset. Optimum cluster size is obtained of each clustering techniques using DB index. The performance analysis of *k*-means, *k*-medoids and fast *k*-means algorithm is calculated using intraclass inertia and select best method which generates compact cluster result. Finally, clusters generated by best clustering method are used to construct portfolio using Markowitz model. Then portfolio comparison is done and selecting outer performing one.

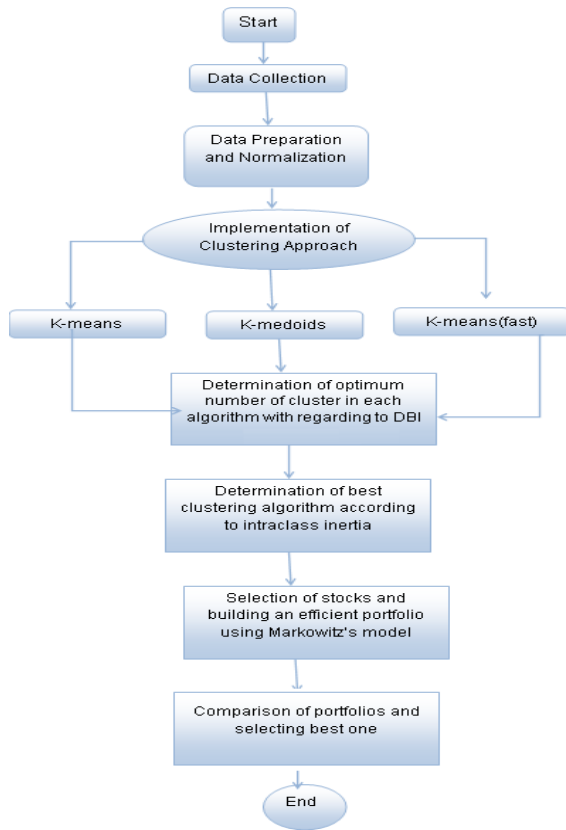


Figure 1. Proposed research methodology

4. DESIGN AND ANALYSIS

The Figure 2 shows an agent based framework for portfolio management.

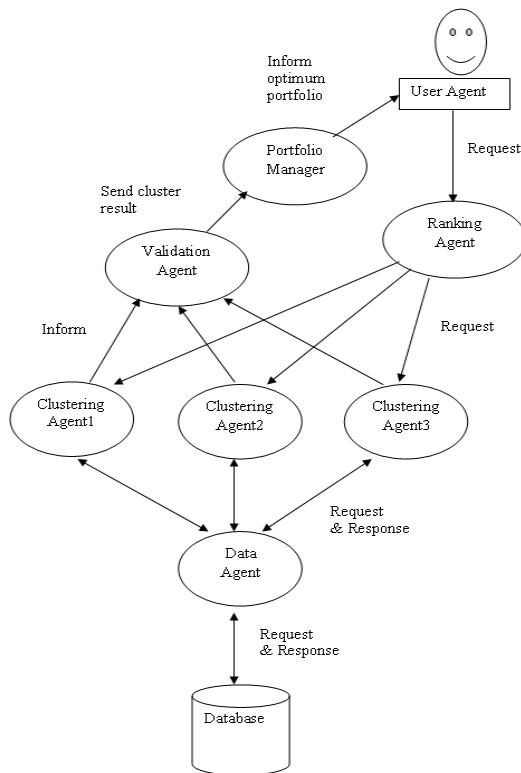


Figure 2. Agent based framework for portfolio management

This framework needs necessary agents to automate portfolio management task. The user agent sets objective that how much assets investors want to invest and delivers result to investor in the form of weighted percentage of each security through user agent. The ranking agent, data agent, three clustering agent and portfolio manager interacts with each other to perform portfolio management task. Ranking agent assigns weights to financial variable so that a data agent collects financial data of Nifty companies. Data agent also collects daily return data of Nifty 50 index to construct portfolio of listed companies. There are three clustering agents used for each clustering algorithm which detects cluster. The clusters generated by each algorithm are validated by validation agent who uses compactness as validation criteria. Validation agent finds optimum cluster size of each clustering algorithm using DB index and performance analysis of each clustering algorithm is calculated using intra-class inertia. The algorithm which generates compact cluster is considered as best clustering algorithm.

5. DATA COLLECTION AND INTERPRETATION

The quantitative data are collected from ACCORD FINTECH financial database which includes list of Nifty 50 stocks' price per earnings ratio, price book ratio, earnings per share, dividend per share and daily return. These stock data of Nifty 50 companies are referred to the start of currency crisis (2012-13) in Indian's economy. The universe of the study is the equity share of company that is listed in the National Stock Exchange Limited, Mumbai, India. For framing of efficient portfolio here the sample of these fifty companies are selected as the Nifty is a well diversified 50 stock index accounting for 23 sectors of the economy. The NSE is located in Mumbai, India and in terms of market capital; it is one of the top 20 largest exchanges in the world. It is used for a variety of purposes such as benchmarking fund portfolios, index based derivatives and index funds. The Nifty index represents about 66.85% of free float market capitalization of the stocks listed on NSE. The Nifty 50 index has also been considered as the market proxy/benchmark index for various studies in literature. Financial ratios and returns of stocks during determined intervals have been presumed variable and they are give in the terms of dimensions of selected stocks. These financial variables are mentioned in Table 1.

Table 1. Financial variables as clustering measures

Financial Variables	Description
Average daily returns for 5 years	Long term returns of company
Earnings per share (EPS)	Net earnings per share made in the business
Dividend per share (DPS)	A portion of firm's profit is distributed to the investors
Price per earnings (P/E)	Eagerness intensity of investors to pay per rupee earning
Book Value per share (BV)	Value of business in terms of net worth per share

6. RESULT AND DISCUSSION

6.1. Optimum Cluster Size

Each algorithm is repeatedly executed by varying cluster size ranging from 2 to 10 and computed DB index as shown in Table 2. It is clear that the smallest absolute DB value of k -means, k -medoids and fast k -means is 0.456, 0.600 and 0.456 respectively. Therefore, an optimum cluster size of k -means, k -medoids and fast k -means is 3, 5 and 3 respectively.

Table 2. DB index for k -means, k -medoids and fast k -means

No. of clusters	Davies-Bouldin Index		
	k -means	k -medoids	fast k -means
2	-0.476	-0.609	-0.476
3	-0.456	-0.850	-0.456
4	-0.550	-0.620	-0.550
5	-0.690	-0.600	-0.690
6	-0.556	-0.779	-0.556
7	-0.801	-0.911	-0.801
8	-0.698	-0.770	-0.698
9	-0.515	-0.706	-0.515
10	-0.869	-0.725	-0.869

6.2. Performance Evaluation

Performance analysis of clustering methods is done using intraclass inertia. Intraclass inertia, $F(k)$ is the measure of compactness of cluster when the number of cluster size is fixed. It can be defined as the average squared Euclidean distance between each observation and its cluster mean. The value of $F(k)$ is linearly scaled in (0, 1) range. From the observation given in Table-3, it is clear that k -means show less variability but stability decrease rapidly in k -medoids and fast k -means by changing value of k .

Table 3. F-measure for k -means, k -medoids and Fast k -means

No. of clusters	F-measure		
	k -means	k -medoids	fast k -means
2	0.607	0.603	0.607
3	0.483	0.764	0.725
4	0.536	0.470	0.598
5	0.575	0.838	1.014
6	0.566	0.538	0.566

7	0.612	0.588	0.627
8	0.520	0.734	0.752
9	0.563	0.818	0.754
10	0.563	0.786	0.820

Table 3 shows F-measure values of k -means, k -medoids and fast k -means for different cluster size ranging from 2 to 10. These $F(k)$ values shows the measure of compactness of clustering methods, are scaled in (0, 1). For example, $F(k)$ measure value for k -means algorithm is 0.607 if number of cluster is 2.

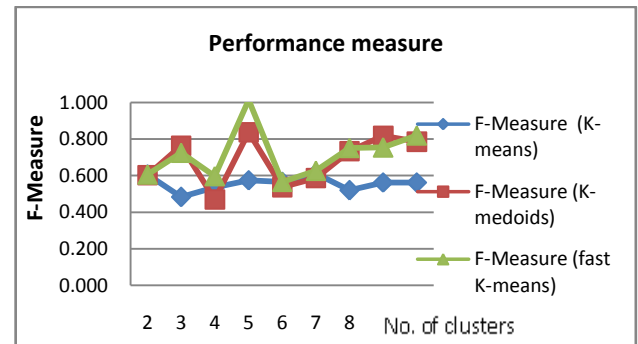


Figure 3. Scaled plot of F-measure for three applied algorithms

The scaled plot of F-measure with respect to different cluster size for each algorithm is given in figure 3. The plot shows that k -means algorithm have less variation as compared to k -medoids and fast k -means. Therefore, a cluster generated by k -means algorithm is considered as final cluster for portfolio construction. Table 4 shows that three clusters are generated having group of 10, 3 and 37 companies.

Table 4. List of companies in three clusters for k -means

Cluster#	#Obs	Companies
Cluster-1	10	ACC Ltd., Axis Bank Ltd., Bank Of Baroda, Dr Reddys Laboratories Ltd., ICICI Bank Ltd., Infosys Ltd., Maruti Suzuki India Ltd., Reliance Industries Ltd., Tata Steel Ltd., Ultratech Cement Ltd.
Cluster-2	3	Grasim Industries Ltd., Punjab National Bank, State Bank Of India
Cluster-3	37	Ambuja Cements Ltd., Asian Paints Ltd., Bajaj Auto Ltd, Bharat Heavy Electricals Ltd., Bharat Petroleum Corporation Ltd., Bharti Airtel Ltd., Cairn India Ltd., Cipla Ltd., Coal India Ltd., DLF Ltd., GAIL (India) Ltd., HCL Technologies Ltd., HDFC Bank Ltd., Hero MotoCorp Ltd., Hindalco Industries Ltd., Hindustan Unilever Ltd., Housing Development Finance Corporation Ltd., IDFC Ltd., IndusInd Bank Ltd., ITC Ltd., Jaiprakash

	Associates Ltd., Jindal Steel & Power Ltd., Kotak Mahindra Bank Ltd., Larsen & Toubro Ltd., Lupin Ltd., Mahindra & Mahindra Ltd., NMDC Ltd., NTPC Ltd., Oil & Natural Gas Corporation Ltd., Power Grid Corporation Of India Ltd., Ranbaxy Laboratories Ltd., Sesa Sterlite Ltd., Sun Pharmaceutical Industries Ltd., Tata Consultancy Services Ltd., Tata Motors Ltd., Tata Power Company Ltd., Wipro Ltd.
--	--

6.3. Portfolio construction

Cluster 1 and 2 consist of 10 and 3 companies for diversification while cluster 3 consists of 37 companies for diversification of portfolio. Cluster 1 and 2 are used for further analysis and Markowitz optimum weights are calculated for efficient portfolios while cluster 3 is ignored for further analysis. As cluster 3 consists of more than 30 stocks for diversification, it is recommended that one can ignore that portfolio for optimization. So what is the optimum level of diversification? This matter has been extensively debated in financial literature for over 30 years; various literatures available [1,2,3,4,6,8,10,12,15,17,32] on the optimum stock size of portfolio. Efficient diversification is critical for controlling risk. However, one has to balance the need for diversification with the time required to monitor the portfolio. The Intelligent Investor, Benjamin Graham (1949) argued that 10 to 30 stocks provide adequate diversification in a portfolio. In literature [11] found that a randomly created portfolio of 32 stocks could reduce the risk by 95%, compared to a portfolio of the entire New York Stock Exchange. It is known as mythical legend that "95% of the benefit of diversification is captured with a 30 stock portfolio." Elton and Gruber (1977) have started the trend for serious thoughts in portfolios. He concluded that most of the benefits from diversification come from adding just the very first few stocks to a portfolio. According to these results adding just 4 more stocks to a 1 stock portfolio gives you 71% of the benefits of diversification in terms of volatility reduction rather adding more and more number of stocks. These results were confirmed in a follow up studies by other researchers in the 1980s and 1990s, the general consensus being that the portfolio of anywhere between 8 and 20 stocks is 'enough'.

6.4. Optimum Markowitz Portfolio

It is a great advantage in terms of reducing the number of possibilities in building an efficient portfolio. This approach can be practically used by investors, specifically at the time of crisis, when all selection of economy such as stock market is influenced tragically. The portfolio is constructed and it is given in Table 5 and Table 6 for cluster-1 and cluster-2. Their weights are calculated by using Markowitz model and concept of risk minimization. This is the minimum variance portfolio.

Table 5. Optimum portfolio for cluster-1

Name of stocks	Weights
ACC	10%
BOB	6%

Dr. Reddys	32%
Infosys	21%
Maruti	21%
Ultratech	18%
Total weight	100%
Mean portfolio target return	37%
Portfolio risk(SD)	18.25%

Table 6. Optimum portfolio for cluster-2

Name of stocks	Weights
Grasim Industries Ltd.	63%
Punjab National Bank	19%
State Bank Of India	18%
Total weight	100%
Mean Portfolio Target Return	18%
Portfolio Risk (SD)	26.29%

From the chart given in Figure 4, one can visualize the efficient frontier of cluster-1 and cluster-2 are drawn where all the efficient portfolios are located. The upper part of the frontier is only feasible solutions for investor as it gives the high chance of return with the increasing level of risk. The lower part of the frontier is inefficient as the returns of the portfolios decreased while the risk of portfolio is increases that can be seen from the level of standard deviation. From the Figure 4, It is clear that cluster-1 with 10 companies dominates the cluster-2 of 3 companies. Investor can have better risk return tradeoff by investing into cluster-1 of 10 companies. As per the above calculation the minimum variance portfolio for cluster of 10 companies would be as follows.

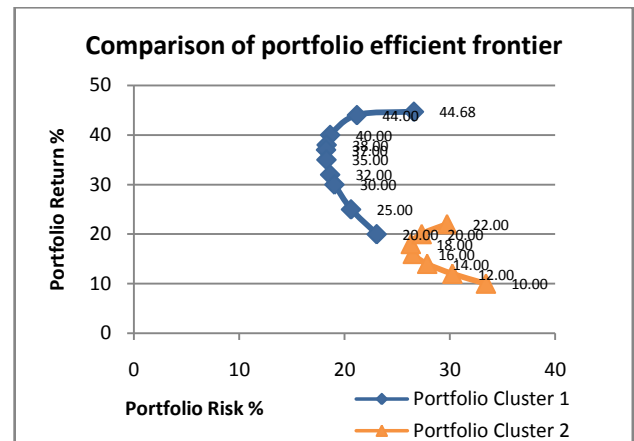


Figure 4. Comparison chart of portfolio efficient frontier

7. CONCLUSION

The present study used financial ratio of Nifty 50 companies to generate cluster by using well known methods named *k*-means, *k*-medoids, and fast *k*-means. Then Davis-Bouldin (DB) index used to find optimum cluster size of each method. The performance of *k*-means, *k*-medoids and fast *k*-means was obtained using intra-class inertia. It was found that *k*-means algorithm generates compact cluster as compared to remaining method. Finally, clusters generated by *k*-means used for investment and portfolio analysis using Markowitz model. This way stock clustering helps investors for framing optimum portfolio and better risk-return profile. When investors spread their investments across different assets in particular proportions, risk of the investment tends to diversified and spread. Thus, this paper can help to assemble a diversified portfolio of stocks with the help of clustering. This paper also helps investor community in specific and in turn it helps the society and economy in general for better allocation of wealth. The research can be further extended by applying other clustering techniques and results can be compared, so that investors can achieve better risk-return profile on their investment.

8. REFERENCES

- [1] A. Kumar and N. William. 2002 "Equity portfolio diversification", Working Paper Yale International Center for Finance, November 2002
- [2] Blume, E. Marshall and F. Irwin, "The asset structure of individual portfolios and some implications for utility functions", *Journal of Finance*, vol 30, 585-603
- [3] C. Bertaut. 1998 "Stockholding behavior of US households: evidence from 1983-1989", *Survey of Consumer Finances, Review of Economics and Statistics*, vol 80, , pp. 263-275
- [4] Cohn , A. Richard et al. 1975 "Individual investor risk aversion and investment portfolio composition", *Journal of Finance*, vol 30, , pp. 605-620
- [5] D. L. Davies and W. Bouldin. 1979 "A cluster separation measure", *IEEE PAMI*, vol 1, pp. 224-227
- [6] E. J. Elton and M. J. Gruber. 1977 "Risk reduction and portfolio size: an analytical solutions", *Journal of Business*, vol 50, , pp. 415-437
- [7] G. Chen et al. 2002. "Evaluation and comparison of clustering algorithms in analyzing ES cell gene expression data", *Statistical Sinica*, vol 12, pp. 241-262
- [8] I. I. Evans and S. H Archer. 1968 "Diversification and reduction of dispersion: an empirical analysis", *Journal of Finance*, vol 41.
- [9] J. C. Dunn, 1974. "Well seperated clusters and optimal fuzzy partitions", *Journal of Cybernatics*, vol 4, pp. 95-104
- [10] J. Heaton and L. Deborah. 2000, "Portfolio choice and asset prices: the importance of entrepreneurial risk", *Journal of Finance*, vol 55, pp. 1163-1198
- [11] L. Fisher and J. H. Lorie, 1970 "Some studies of the variability of returns on investment in common stocks", *Journal of Finance*, vol 43.
- [12] L. Guiso, T. Japelli and D. Terlizze, 1996. "Income risk, borrowing constraints and portfolio choice, *American Economic Review*", vol 86, pp. 158-172
- [13] L. Kaufma and P. J. Rousseeuw, 1990. "Finding groups in data: and introduction to cluster analysis", John Wiley & Sons.
- [14] M. Halkidi, Y. Batistakis and M. Vazirgiannis, 2002. "Cluster validity methods part II", *SIGMOD Rec*, vol 31, issue 2, pp. 40-45
- [15] M. Kelly, "All their eggs in one basket: portfolio diversification of US households", *Journal of Economic Behavior and Organization*, vol 27, pp. 87-96
- [16] M. Kim and R. S. Ramakrishna, 2005. "New indices for cluster validity assessment", *Pattern Recognition Letters*, vol 26, issue 15, pp. 2353-2363
- [17] M. Satman, 2002. "How much diversification is enough?", Working Paper, Santa Clara University August 2002
- [18] N. Malhotra, 2013. "Determinants of stock prices: Empirical evidence from NSE 100 companies", *Int J Res Mangg Technol*, vol 3, issue 3, pp. 86-95
- [19] P. Baser and J. R. Saini, 2013. "An agglomerative analysis of Nifty companies for an investment perspective", *Int J Recent Innov Trends Comput Commun*, vol 2 issue 5, pp.1348-1352.
- [20] P. Baser and J. R. Saini, 2013. "A comparative analysis of various clustering techniques used very large datasets", *Int J Comput Sci Commun Netw* vol 3, issue 4, pp. 271-275
- [21] P. Baser and J. R. Saini, 2013. "An intelligent agent based framework for an efficient portfolio management using stock clustering", *Int J Inform Comput Technol*, vol 3, issue 2, pp. 49-54.
- [22] P. Baser and J. R. Saini. 2014 "An Optimum Cluster Size Identification for k-Means using Validity Index for Stock Market Data", *Int J Data Mining Emer Technol*, vol 4, issue 2, , pp. 107-110
- [23] P. Baser and J.R. Saini. 2013 "k-Means analysis of Nifty companies for an investment perspective" *Int J Data Mining Emer Technol*, vol 3 issue 1, pp.16-22.
- [24] P. S. Nirmal, P.S. Sanju and M. Ramachandran. 2011, "Determinants of share prices in india" *J. Emer Trend Econ Mangg Sci*, vol 2, issue 2, pp. 124-130
- [25] P. Srinivasan, 2011. "Determinant of equity share prices in india : A panel data approach", *Romanian Econ J Mangg Sci*, vol 2, issue 2, pp. 124-130
- [26] S. K. Mehta and M. S. Turan, 2005. "Determinants of stock prices in india : An empirical study", *J Ind Mangg Strat*, vol 10, issue 4, pp. 37-43.
- [27] S. Kambey, R.Thakur and S. Jaalori, 2012. "Application of data mining in stock market: an analysis", *International Journal of Computer and Communication Technology(IJCCT)*, vol 3, issue 3, pp.52-53.
- [28] S. Kaur and V. Mangat, 2012. "Applications of data mining in stock market". *Journal of Information and Operations Management* vol 3, issue 1, pp. 86-88.
- [29] S. Sharma and B. Singh, 2006. "Determinants of equity share prices in indian corporate sector: an empirical study" *ICFAI J Appl Fin* , vol 12, issue 4, pp. 21-28.

- [30] S. Sharma, 2011. "Determinants of equity share prices in india", J Arts Sci Comm, vol 2, issue 4, pp. 51-60
- [31] U. Malik and S. Bandopadhyay. 2002. "Performance evaluation of some clustering algorithms and validity indices" IEEE Transc and Pattern Analysis and Machine Intelligence, vol 24, issue 12, , pp. 1650-1654
- [32] W. R Perraudin and E. Bent, 2000. "The demand for risky assets: Sample selection and household portfolios", Journal of Econometrics , vol 97, pp. 117-144
- [33] W.C. Elkan 2003. "Using the Triangle Inequality to Accelerate k -Means" Proceedings of Twentieth International Conference on Machine Learning, Washington DC,(ICML-2003).