ISSN 0971-0388

Aligarh Journal of Statistics
Vol. 41 (2021*)*, 63-94

Taxonomical Grouping of Firms: A Study on Listed Banks in India

Naseem Ahamed

[Received on January, 2020. Accepted on March, 2021]

ABSTRACT

The objective of this article is to categorize homogeneous stocks using cluster analysis methodology. It is easier for investors to deal with a small number of clusters of stocks compared to dealing with thousands of stocks. This has aroused their interest in comparing stocks with respect to different variables based on their risk appetite. Notwithstanding, the widespread availability of high processing power computing devices, investors get overwhelmed by the large number of stocks available at their disposal. Categorizing the large number of stocks into few distinct clusters would not only make the task easier for investors by letting him deal with less number of data, but would also give him the option to pick stocks from different clusters based on his preference. This article uses the cluster analysis methodology to group homogeneous stocks from a dataset of 33 listed Indian banks. This method provides a useful tool for interpolation and extrapolation of statistical data and sets up a measure to compare performance and profitability of a company.

1. Introduction

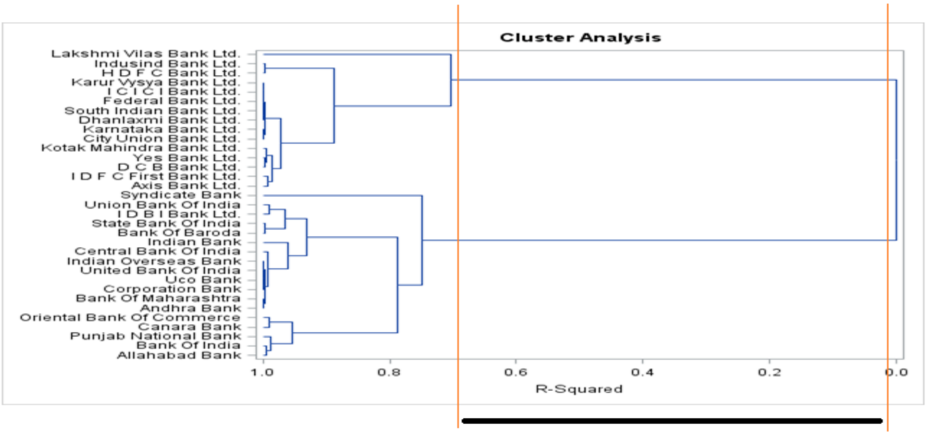
Investors/Portfolio managers are better-off when they take informed decision. They have to decide to include a stock in their portfolio based on its risk-return characteristics; past performance; financial parameters of the stocks etc. Finally, the investor would choose a couple of stocks from the universe of stocks of a particular sector. Hence, it would be worthwhile for him if all the stocks of a particular sector can be categorized into few distinct groups. This article focuses on the stocks of the banking sector of India. The banking sector is one of the most important sectors of any economy in general and developing economy in particular. An emerging economy, unlike its advanced counterparts doesn’t have strong and robust institutions; lacks managerial talent and faces the consequences of regulatory shortcomings/lack of regulatory implementation. It leads to an institutional void in developing economy where access to markets is limited for interested participants.

✉ : Department of Finance and Accounting, IBS Hyderabad, IFHE University, Hyderabad, India-501203.
E-mail: naseemahamed@ibsindia.org

Banks serve as an alternative to the capital market in developing economies, where they not only provide capital but also render several other useful services. *Inter alia* banks pool funds from depositors and channelize them into economically beneficial projects. State owned banks lend to sectors prioritized by the government such as agriculture; cottage industry etc. in order to uplift those sectors. Banks fuel the growth of a developing economy where access to market is limited. India has made significant strides in expanding and modernizing its security market but there is still room for improvement. Banks have played a key role in the development of the Indian economy. Investors looking to include a few banking stocks in their portfolio can apply the taxonomical clustering approach.

Making a brief profile about each potential stock[2] and selecting some from them could get overwhelming for the investor. Grouping the stocks of similar characteristics together can reduce this problem. These clusters would be far less in number that the original pool of stocks. Another advantage of this technique is that it can also suggest the number of groups that can be formed based on the data distribution after calculating their distances from each other. A clustering technique is used to categorize entities based on similarity of variables. It groups entities into clusters in such a way that there is more homogeneity in the cluster than between the clusters.

Figure 1: Number of clusters.



Note: In figure 1, the method of finding the suitable number of clusters through hierarchical clustering method is exhibited. The best estimate of the number of clusters is to find the maximum distance that can be secured without transgressing any other union in a dendrogram tree.

[2] The number of such stocks could run into hundreds or more.

The clustering analysis methodology represents a battery of techniques that are used to group similar elements together and separate dissimilar elements. One of the most popular methods of clustering is the hierarchical method[3], which is classified into agglomerative approach and divisive approach respectively. Both these approaches work opposite of each other where the agglomerative approach starts by considering the each element of the dataset as a cluster in and of itself and then starts adding more elements into each cluster based on their vicinity through distance measure. The divisive approach starts out by considering the entire dataset as one single cluster and then starts separating distant elements from the cluster giving rise to another cluster.

One of the ways to find the best estimate of the number of clusters is to observe a sharp kink in the Cubic Clustering Criterion (CCC). Another way of finding the best estimate of the number of clusters is to find the maximum distance that can be secured without transgressing any other union in a dendrogram tree. It is represented in figure 1 where we can see the longest distance in the dendrogram tree is for two unions of cluster.

Another important clustering method is known as the K means clustering in which a random number of clusters are assigned to the dataset. The clusters are then assigned centroids closer to the elements of assigned clusters. The elements closest to the centroid are assigned clusters closest to them and this process repeats itself until a better solution is reached. Finally, the method would eventually come up with the best possible number of clusters to classify the elements of the dataset. There are different measures to measure the distance between elements of a dataset for grouping them into clusters. Using either of these distance measures[4] the clustering technique classifies elements into few clusters.

[3] In the hierarchical method of clustering, the principle behind two elements inked to each other is the distance between them. Objects with small distance would be part of a cluster and objects far away would be part of another cluster. Hence, at different distances, different clusters would be formed which in turn would create a series of hierarchical clusters displayed in the form of a dendrogram. Other important clustering techniques consists of centroid based clustering, density based clustering etc.

[4] The different measure of distance are as follows:

Euclidean distance: $\|a-b\|_2 = \sqrt{\left(\Sigma\left(a_i-b_i\right)\right)}$;

Squared Euclidean distance: $\|a-b\|_2^2 = \Sigma\left(\left(a_i-b_i\right)^2\right)$;

Manhattan distance: $\|a-b\|_1 = \Sigma\left|a_i-b_i\right|$;

Maximum distance: $\|a-b\|_{\text{inf}} = \max_i \left|a_i-b_i\right|$;

This article uses cluster analysis to categorize the listed banks in India based on ten variables[5]. The banking industry forms a significant chunk of the total market and many investors want bank stocks in their portfolios. Grouping banks on the basis of the parameters selected by the investor makes the task significantly easier for him. He can design the framework for investment criteria based on performance variables. The remaining of the article is outlined as follows: Section 2 presents the literature review. Section 3 contains the theoretical framework in detail. Section 4 discusses the data and methodology being followed and the various indicators used. Section 5 describes the results and findings of the research. Section 6 concludes the paper. Section 7 highlights the limitations and scope for further studies.

2. Literature Review

The categorization of homogeneous entities together using the taxonomical technique of clustering has been used extensively in literature (see Cormack, 1971). In recent times, the usage of the clustering technique has gained traction in the field of management. Categorization by clustering group elements after identifying similarities among different elements is the basis of the technique. Jensen (1969) proposed using a dynamic programming algorithm to deal with complex situations and arrive at the best clustering solution. A dynamic algorithm updates the existing clusters on the basis of latest inputs received after recalculating the distances among the elements of the dataset. Although the theoretical foundation was present, the lack of intensive computing devices were a major limitation in executing such model.

Using a similar technique, Bensmail and DeGennaro (2004) applied a new and robust statistical modelling technique to cluster analysis on the financial data for Federal Reserve Bank of Atlanta. The authors served a two pronged purpose with the techniques. First, they handled the issue of missing data from the dataset and they also found homogeneous group within the dataset. Meyers (1973) analyzed the mechanism through which market and industry factors are absorbed into the stock prices and the subsequent fluctuation of the stock price. It focused on the systematic risk that can't be diversified away with increase in the number of stocks in the portfolio. The author asserts that the role of industry effects were overstated by King (1966) because he found the industry effect much less than expected. His finding corroborates that grouping of stocks on factors other than industry parameters would

Mahalanobis distance: $\sqrt{(a-b)^T (S)^{-1} (-b)}$ (where, S is the covariance matrix)

[5] The name of all the variables; their calculation and data source is given in section 4 of this article.

help determine stock price variation. Miceli and Susinno (2003, 2004) used the cluster analysis technique to categorize hedge funds on the basis of returns generated. The authors favour the usage of clusters for the ease of interpretation when compared to large correlation matrices. The clustering can be achieved in an ‘n’ dimensional vector space. Hence, the authors assert that selection of stocks from different clusters would lead to a better diversified portfolio. They argue that the tree structure provides an objective basis for extraction of economic conclusions parsimoniously.

Martin (2001) found the existence of considerable heterogeneity in constituent funds within the clusters in his study on monthly returns generated by hedge funds. The structure of principal component analysis (PCA) on returns generated by several funds is found to be different from that of a standard equity of hedge fund index (see Kazemi, Gupta and Daglioglu (2003). Das (2003) concludes that the results of cluster analysis are more robust than the ZCM/Hedge fund classifications in grouping historical managerial returns after classifying managers based on asset class, style of hedge fund, incentive fee, risk level, and liquidity. Gibson and Gyger (2007) use cluster analysis in their study to conclude that managers are not consistent with their investment style consistently over time. Fuzzy clustering is used to illustrate the degree of misclassification existing in the industry-accepted investment-style classifications. Haldar *et al.*, (2008), also use cluster analysis in order to group patients with asthma exhibiting clinically relevant differences in outcome for titrating corticosteroid therapy. They performed k-means cluster analysis in three independent asthma populations. The dataset were clustered at entry into a randomized trial comparing a strategy of minimizing eosinophilic inflammation with standard care. Cluster analysis provided a novel multidimensional approach for identifying asthma phenotypes that exhibit differences in clinical response to treatment algorithms.

Alexandra *et al*., (2008) used cluster analysis to classify the financial performance of firms in Central and Eastern Europe. Ahamed and Bhattacharjee (2012) classified the fixed income securities market with the assertion that equity market is much extensively studied when compared to fixed income security market. Jain and Subramanyam (2015) used the taxonomical classification of companies in the IT sector based on ten different financial and non-financial variables.

In this article, cluster analysis technique is employed to categorize together similar stocks of banking companies listed on the National Stock exchange (NSE). It is expected that the stock performance of companies based in same sector would be similar due to restrained macroeconomic environmental variables. The variations can arise due to microeconomic variables.

3. Theoretical Framework

Cluster analysis is a non-parametric statistical tool for categorizing entities similar to one another. It measures the distance among various entities represented as data points for classification. Data points with less distance with respect to one another are grouped together and data points farther away are grouped together. The clustering method has its roots in taxonomy known as Wroclaw taxonomy that was designed by Polish mathematicians in order to obtain a statistical method of determining homogeneous units or 'types of things' in an n-dimensional vector space, without the use of regression, variance or correlation analysis.

This article used hierarchal clustering that involves the calculation of distance matrix which is already done in the taxonomic method. Distance is a measure of how far apart two objects are, while similarity measures how similar two objects are. For cases that are alike, distance measures are small. On the basis of the distance matrix, the data is analyzed and clustered.

The taxonomic method is illustrated to appreciate the theoretical underpinning and application of the method. Let's assume that a set of n points representing units 1, 2,…, n for a group of variables 1, 2,...., m represented by the following matrix:

Matrix 1

$$\begin{pmatrix} X_{11} & X_{12\cdots} & X_{1m} \\ X_{21} & X_{22\cdots} & X_{2m} \\ . & . & . \\ . & . & . \\ X_{n1} & X_{n2\cdots} & X_{nm} \end{pmatrix}$$

Hence, each unit is represented by a vector in an n-dimensional space. Normalization of variables is executed by the following formula:

$$\frac{X_i - \overline{X_j}}{\sigma_j} \quad \text{where} \quad j = 1, 2, \ldots, m$$

In order to find the values of $\overline{X_j}$ and σ_j, we apply the following equations:

$$\overline{X_j} = \frac{1}{N}\Sigma_{i=1}^{N} X_{ij} \quad \text{and} \quad \sigma_j = \sqrt{\left[\frac{1}{N}\sum_{i=1}^{N}\left(X_{ij} - \overline{X_j}\right)^2\right]}$$

The normalization of the dataset results in a new matrix, where each unit is represented by a standardized vector in an m-dimensional space. The normalized matrix would be as below.

Matrix 2

$$\begin{pmatrix} D_{11} & D_{12}\cdots & D_{1m} \\ D_{21} & D_{22}\cdots & D_{2m} \\ . & . & . \\ . & . & . \\ D_{n1} & D_{n2}\cdots & D_{nm} \end{pmatrix}$$

Where $D_{11} = \frac{X_{11}-\overline{X_1}}{\sigma_1}$; $D_{12} = \frac{X_{12}-\overline{X_2}}{\sigma_2}$; $D_{1m} = \frac{X_{1m}-\overline{X_m}}{\sigma_m}$

After obtaining the standardized matrix i.e. matrix 2, we have to find the difference from a point to every other point (1, 2,…, n) for each of the m variables, which results in matrix 3.

Matrix 3

$$\begin{pmatrix} D_{11}-D_{21} & D_{12}-D_{22}\cdots & D_{1m}-D_{2m} \\ D_{11}-D_{31} & D_{12}-D_{32}\cdots & D_{1m}-D_{3m} \\ . & . & . \\ . & . & . \\ D_{n-1}-D_{n1} & D_{(n-1)2}-D_{n2}\cdots & D_{(n-1)m}-D_{nm} \end{pmatrix}$$

The difference between any two points P_a and P_b for any set of m variables is derived by the following formula:

$$c_{ab} = \sqrt{\left[\sum_{k=1}^{m}\left(D_{ak}-D_{bk}\right)^2\right]}$$

Where $c_{aa} = 0;\ c_{ab} = c_{ba};\ c_{ab} \le c_{ak} + c_{kb}$

The equation above results in a symmetric matrix termed as the distance matrix:

Matrix 4

$$d_{ij}=\sqrt{\sum_{i=1}^{p}\frac{\left(X_{i_i}-X_{i\,j}\right)^2}{p}}$$

After obtaining matrix 4, the minimum distance from a given unit to all other units in the row can be found that is the index of resemblance[6]. The next step is to determine

[6] The closest point in a given frame of reference.

the critical region[7] between which the minimum distance found is considered as significant. Finally find the ideal value for each variable in every set of n.

4. Data and Methodology

The methodology adopted by this study is clustering analysis to categorize similar stocks on the basis of different variables. Clustering is a useful because it reduces the complexity of a population into manageable macro classes. The theoretical framework of the methodology used in this study is provided in section 3.

a) Data Source

A set of 40 banking companies are taken for the study from the Prowess database (a comprehensive database on Indian economy maintained by the Centre for Monitoring Indian Economy). Seven banks are removed from the dataset because of either non-availability of data or merger of banks. Hence, the number of banks for the final study is 33 (Indicated in Appendix). The companies belong to banking sector and are shown in the appendix section. A total of 10 variables are taken on which the analysis will be carried out. The variables considered for this study are mentioned below:

b) Variable Construction

The following variables are used in the study and they are constructed as below:

i) **Promoters Holding:** The proportion of ownership by the promoters of the company is measured through this variable. Multiple studies have shown the impact of promoter’s holding on financial performance of firm where a higher degree of promoter ownership has been associated with both better management and tunnelling wealth.

ii) **Government of India Holding:** The ownership proportion of the Government of India over the company is measured through this variable. Several studies in the stream of corporate governance and executive turnover indicate that state

[7] The critical minimum distance is derived by the formula $c(+) = \bar{c} + 2\sigma$ Where $\bar{c} = \frac{1}{N} \sum_{j=1}^{N} c_j$ is the arithmetic mean of the distances c_j, the minimum in each row of the distance matrix and $\sigma = \left[\frac{1}{N} \sum_{j=1}^{N} \left(c_j - \bar{c} \right)^2 \right]^{0.5}$ is the standard deviation of the minimum distances in each row. The number of n elements in the set can be reduced further with the second critical value $c(-) = \bar{c} - 2\sigma$

The critical value may be considered as a measure of resemblance; the greater is c (+), the smaller is the resemblance between all possible pairs of points.

owned firms don't take performance into account when making policy decisions. Government can prioritize other goals such as welfare of a sector; upliftment of a backward region etc. Hence, ownership by the state plays an important role in corporate decisions. Banks with a large government ownership would be generally safer than other private banks but it wouldn't be as profitable as their private counterparts.

iii) **Debt to Equity Ratio:** This variable measures the ratio of debt to equity which is an important consideration for any company. Firms get the benefit of tax shield on interest payments of debt encouraging them to take debt. On the other hand, burdening the capital structure with too much debt would increase the bankruptcy risk.

iv) **Beta:** The systematic risk or market based risk or undiversifiable risk is measured by beta. This risk component is non-diversifiable with increase in number of stocks in the portfolio. It represents the sensitivity of the stock with respect to the market. A stock with a beta value higher than one is aggressive than the market and *vice-versa*.

v) **Price to Book Ratio:** This variable is called price to book ratio and it is calculated by dividing the market price of the stock with its book value. A high value of price to book ratio indicates that the stock is overvalued and *vice-versa*.

vi) **EPS:** This variable stands for earnings per share (EPS). It is calculated by dividing total profit after taxes by total number of shares outstanding. It is an important indicator for the shareholders as it is a direct measure of the company's profitability.

vii) **Current Ratio:** This variable demonstrates the company's ability to honour its short term obligations. It is calculated by dividing the current assets (cash, cash equivalents, marketable securities, receivables and inventory) by the current liabilities (term debt, payables, accrued expenses and taxes). A higher current ratio is considered better for the company but too high a current ratio indicates that the assets of the company are not being properly utilized.

viii) **Return on Assets:** This variable demonstrates the managerial ability to efficiently utilize the assets of the company. It is calculated by dividing the earnings before taxes by the total assets. A higher return on assets is considered better for the company.

ix) **Return:** The return represents the stock return calculated from closing prices of stocks. These prices are corrected for dividends, splits, and other events.

$$\mathrm{Re}\,turn = \left[\frac{P_t - P(t-1)}{P(t-1)}\right]$$

Where P_t is the closing price of stocks at time t.

x) **Risk:** The total risk represents the deviation from the expected return (measured by mean). It is measured by standard deviation bearing the sign sigma 'σ'

$$\sigma = \left[\frac{\Sigma\left(r_i - \bar{r}\right)^2}{n-1}\right]^{0.5}$$

Where r_i is the nominal return; $\bar{r}$ is mean return. Variables are standardized to prevent units from interfering with the weights of individual variables.

$$SV_i = \frac{\left(X_i - \overline{X_1}\right)}{\sigma}$$

Now, Euclidean Distance between stocks is represented by d_{ij}.

$$d_{ij} = \sqrt{\sum_{i=1}^{p} \frac{\left(X_{i_i} - X_{i_j}\right)^2}{p}}$$

Here, X_{i_i} gives the location of stock i compared to plane i's origin, and p is space size, i.e. the number of variables.

5. Results and Findings

The study uses both hierarchical clustering and k-means cluster algorithm of Aldenderfer and Blashfield (1984). We use the statistical software SAS 9.4 for running the models of this study. The descriptive statistics of all the variables for both methods used in this study is provided in table 1. The mean promoter holding is more than 45 percent in the banks listed in India which establishes concentrated ownership pattern in the Indian corporate landscape. The mean age of the banking firms is almost 78 years indicating that the banks are fairly old. They have witnessed the evolution of banking business through different political and technological changes with the passage of time. Whereas, many large banks operating in the country can be referred to as legacy banks that have large fixed cost operational expenses, other emerging new age banks rely heavily on digital infrastructure cutting down their operational expense. Most of the public deposits still find their way into the old legacy banks because people trust them more. The mean beta value is 1.6 alluding that the aggressive movement of stocks with respect to market. The values of other variables are exhibited in table 1.

The variables are measured in different units and differ in magnitude. In order to make the variance of the variables equal, they have been standardized where their mean becomes 0 and standard deviation becomes 1. The standardized values of the variables that would be used for both hierarchical and k means clustering method are shown in table 2.

Suspecting the presence of elliptical shaped cluster, the data is transformed so that the within cluster covariance matrix is spherical. Approximate estimates of the pooled within cluster matrix covariance matrix is computed and then canonical variables are computed to be used in subsequent analyses. This study analyzes the dataset in two ways. First, hierarchical clustering is executed for all companies in the dataset using 10 variables namely Promoter's holding, Government of India holding, Debt to equity ratio, beta, price to book ratio, Earnings per share, Current ratio, Return on assets, return generated and risk. Second, the k-means clustering is executed for all companies in the dataset using all the 10 above mentioned variables.

a. Hierarchical Clustering Using Ten Variables

The Eigen values of the covariance matrix that is used in the computation of Cubic Clustering Criterion (CCC) is exhibited is table 3. Eigen value is the value of an Eigen vector which does not rotate the vector but expands/contracts. The CCC is used as one of the criteria to determine the number of clusters that can be formed. It can be observed from table 3 that 99.9 percent of the variation associated with Eigen vales can be explained by 5 clusters only.

Table 3: Eigen value of covariance matrix.

	Eigen value	**Difference**	**Proportion**	**Cumulative**
1	3607.58274	2709.20881	0.7712	0.7712
2	898.37392	731.87825	0.1920	0.9632
3	166.49568	162.56155	0.0356	0.9988
4	3.93412	2.95983	0.0008	0.9997
5	0.97429	0.42834	0.0002	0.9999

	Eigen value	**Difference**	**Proportion**	**Cumulative**
6	0.54595	0.47979	0.0001	1.0000
7	0.06616	0.06610	0.0000	1.0000
8	0.00006	0.00003	0.0000	1.0000
9	0.00004	0.00003	0.0000	1.0000
10	0.00000		0.0000	1.0000

Root mean square Total sample Standard Deviation 21.62
Root mean square Distance between observations 96.72

Note: Table 3 above exhibits Eigen values of covariance matrix. The first column represents each Eigen value. The second column represents the difference between the Eigen value and its successor. The third column exhibits the proportion of variance associated with the corresponding Eigen value. The last columns exhibit the cumulative proportion of variance associated with each Eigen value.

The history of cluster generation is shown in Table 4 where it can be observed that the variance explained by the clusters exceed 85 percent, when the number of clusters are 5. The values in Cubic Clustering Criterion (CCC), Pseudo F statistic and Pseudo t^2 help in determining the suitable number of clusters in which the elements of the dataset are grouped. A sharp kink in the value of CCC indicates the number of clusters. In Table 4, we observe local peak/sharp changes in the value of CCC at cluster number 2 (The value of CCC changes from 0.00 to 2.24) and cluster number 5 (The value of CCC changes from -0.87 to 1.78).

Similarly, observing the values of Pseudo t^2, it can be found sharp changes in the value of Pseudo t^2 at cluster number 2 (The value of Pseudo t^2 changes from 73.4 to 9.5) and cluster number 5 (The value of Pseudo t^2 changes from 30.2 to 25.3). Hence, we would take five clusters for our dataset.

Table 4: Cluster History.

Cluster	**Cluster Joined**		**Freq**	**Semi partial R^2**	**R^2**	**$ExpR^2$**	**CCC**	**Pseudo F Statistic**	**Pseudo t^2**
15	Axis Bank Ltd.	I D F C First Bank Ltd.	2	0.0009	0.992	.	.	162	.
14	Canara Bank	Oriental Bank Of Commerce	2	0.0011	0.991	.	.	163	.

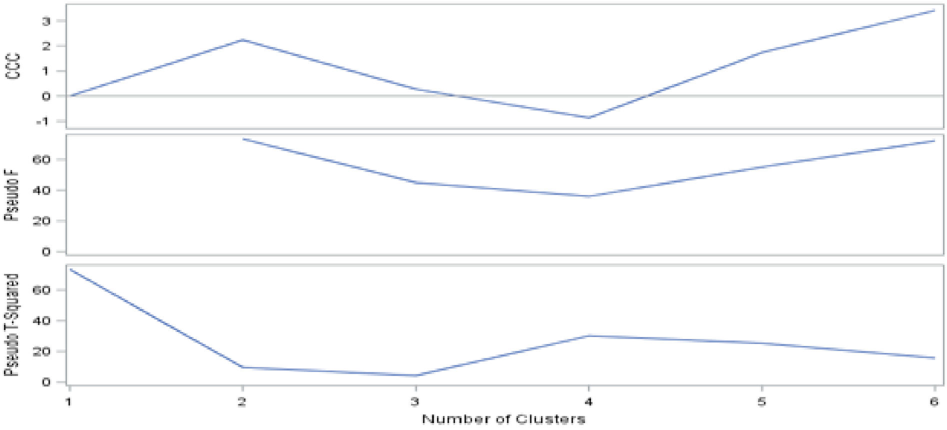
Cluster	Cluster Joined		Freq	Semi partial R^2	R^2	ExpR^2	CCC	Pseudo F Statistic	Pseudo t^2
13	I D B I Bank Ltd.	Union Bank Of India	2	0.0012	0.990	.	.	163	.
12	CL18	Punjab National Bank	3	0.0019	0.988	.	.	157	3.2
11	CL15	CL17	5	0.0034	0.985	.	.	141	4.3
10	CL11	CL23	12	0.0133	0.971	.	.	86.7	19.4
9	CL19	CL13	4	0.0053	0.966	.	.	85.5	6.3
8	CL16	Indian Bank	8	0.0054	0.961	.	.	87.4	10.9
7	CL12	CL14	5	0.0065	0.954	.	.	90.3	5.6
6	CL8	CL9	12	0.0238	0.930	0.884	3.42	72.2	15.6
5	CL10	CL22	14	0.0433	0.887	0.855	1.74	55.0	25.3
4	CL7	CL6	17	0.0986	0.789	0.812	-.87	36.0	30.2
3	CL4	Syndicate Bank	18	0.0388	0.750	0.741	0.28	44.9	4.2
2	CL5	Lakshmi Vilas Bank Ltd.	15	0.0466	0.703	0.600	2.24	73.4	9.5
1	CL3	CL2	33	0.7031	.000	.000	0.00	.	73.4

Note: Table 4 above exhibits the previous 15 generations of cluster history. The column semi-partial r^2 value represents the decrease in the proportion of variance accounted for by joining the two clusters. The r^2 value in the next column displays the proportion of variance accounted for by the clusters. The last three columns display the values of the cubic clustering criterion (CCC), pseudo F (PSF), and Pseudo t^2 (PST2) statistics. These statistics are useful for estimating the number of clusters in the data.

The two dimensional contour of CCC, Pseudo F statistic and Pseudo t^2 of table 4 is presented graphically in figure 2.

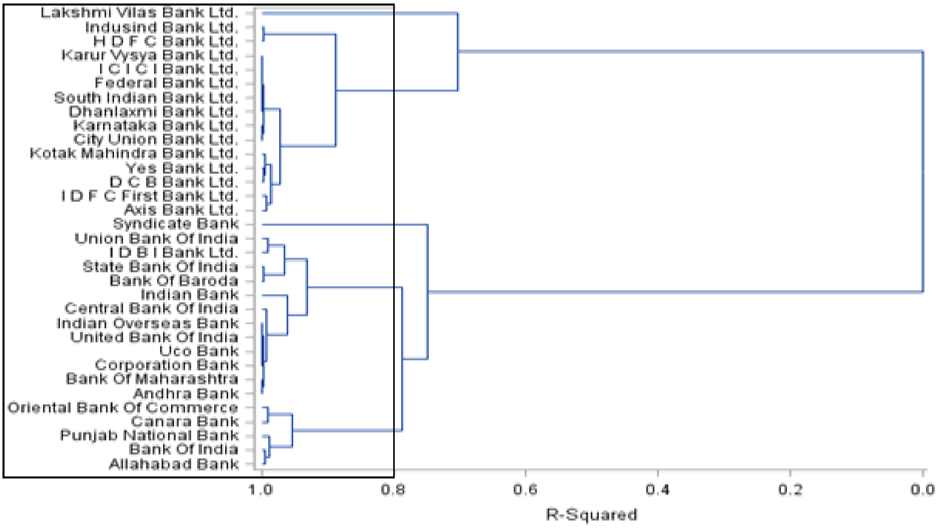
Note: Figure 2 shows the criteria for the number of cluster formation is exhibited. Sharp kinks in the value of CCC are used for estimating the number of clusters. In the figure, there is a local peak of the CCC i.e. a sharp decline forming a kink when the number of clusters is five.

Figure 2: Cluster formation criteria.



Observing the value of CCC from right to left, we can find a sharp decline when number of clusters are 5 and when number of clusters are 2. The dendrogram shown in figure 3 displays the distance at which different stocks are grouped in a cluster. It can be observed that as the number of branches grows to the left from the root, the value of r^2 approaches 1; the first five clusters (branches of the tree) account for more than half of the variation (about 80 percent). In other words, only four clusters are necessary to explain over 80 percent of the variation.

Figure 3: Dendrogram of clusters versus r^2 values.

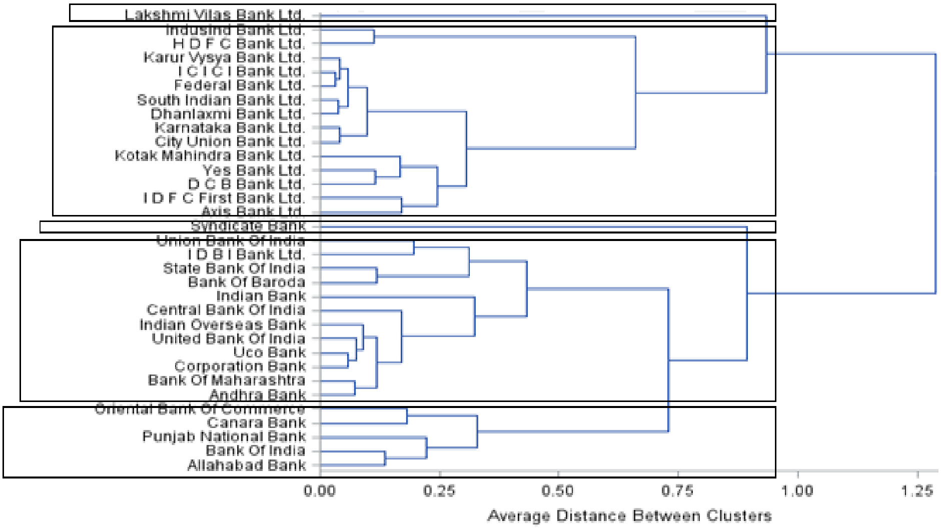


Note: Figure 3 exhibits a dendrogram which provides a graphical view of the information in Figure 2. As the number of branches grows to the left from the root, the value of r^2 approaches 1; the first five clusters (branches of the tree) account for over half of the variation (about 80%). In other words, only four clusters are necessary to explain over 80% of the variation.

The clustering process of elements in the dataset with other elements or elements with sub-clusters joining at different distances is exhibited in Figure 4. It can be clearly observed that four distinct clusters are formed using the average distance measuring method. Other distance measuring methods would give results with slight variations.

The correlation coefficients amongst the stocks are used in the form of inputs in a likelihood matrix and stocks are then merged based on similarity. The resulting clusters are exhibited through dendrograms representing hierarchical organization between stocks. The results of our study indicate that stocks can be clustered based on their vicinity and thereby parsimoniously utilized for the purpose of decision making by the investors or fund managers. Every stock in the dataset is considered to have equivalent weight in a cluster. Once the best cluster is determined, the investor can make his choice by selecting a stock within the cluster.

Figure 4: Dendrogram of clusters with average distance between clusters.



Note: Figure 4 exhibits a dendrogram which provides a graphical view of the information in Figure 2. The dendrogram exhibits the average distance between clusters where two data points or sub-clusters join. The figure above clearly exhibits the existence of five distinct clusters.

Under the assumption that the distribution of data is uniform, there are five categories of banks based on the ten variables. The elements are converted so that the within cluster variance matrix is spherical. Then the distance among the elements are measured to form clusters. If multiple sectors are considered then the investor can choose stocks considering the industry in which the company operates. In Table 5, the average distance measuring method is used to categorize companies.

Table 5: Names of companies in different clusters (Clustering method = Average).

Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Column 5**
Federal Bank Ltd.	Corporation Bank	Allahabad Bank	Syndicate Bank	Lakshmi Vilas Bank Ltd.
I C I C I Bank Ltd.	Uco Bank	Bank of India		
Dhanlaxmi Bank Ltd.	Andhra Bank	Canara Bank		
South Indian Bank Ltd.	Bank Of Maharashtra	Oriental Bank of Commerce		
Karur Vysya Bank Ltd.	United Bank Of India	Punjab National bank		
City Union Bank Ltd.	Indian Overseas Bank			
Karnataka Bank Ltd.	Bank Of Baroda			
H D F C Bank Ltd.	State Bank Of India			
Indusind Bank Ltd.	Central Bank Of India			
D C B Bank Ltd.	I D B I Bank Ltd.			
Yes Bank Ltd.	Union Bank Of India			
Kotak Mahindra Bank Ltd.	Indian Bank			
Axis Bank Ltd.				
I D F C First Bank Ltd.				

Note: Table 5 exhibits the categorization of banking companies in our dataset using the average distance measuring method between clusters.

Using a different distance measuring method might result in different number of clusters in some cases. However, largely the clusters and elements therein remain the same. Other cluster tables are not exhibited in this article as they are by and large similar to the one provided in table 5.

As a measure of robustness check, the mean return of stock samples in a cluster is compared with that of another cluster. In Table 6 the t coefficients of a two sample t test is shown. The results of this test indicate that the mean values of stocks belonging to different clusters are different. This establishes that the clustering process has grouped dissimilar companies in different clusters. The results of table 6 clearly indicates that stocks of banking firms placed in cluster 1 are statistically different from those placed in cluster 2. Further, the stocks of cluster 1 generate higher returns when compared to the stocks of cluster 2. The results of t test for other clusters are given in table 10 and 11 in the appendix section at the end.

Table 6: t test table for cluster 1 and 2.

	N	**Mean**	**Std Dev**	**Std Error Mean**
Cluster 1	14	0.003	0.0026	0.001
Cluster 2	12	-0.002	0.0028	0.001
Observed difference (Cluster 1 - Cluster 2)	0.005			
Standard deviation of difference	0.0011			
Unequal Variances				
Degree of freedom	22			
95% Confidence interval for the difference	0.0027; 0.0073			
t test statistic	4.54			
Cluster 1 ≠ Cluster 2 (p value)	0.00			
Cluster 1 < Cluster 2 (p value)	0.99			
Cluster 1 > Cluster 2 (p value)	0.00			
Equal Variances				
Degree of freedom	24			
95% Confidence interval for the difference	0.0027; 0.0073			
t test statistic	4.70			
Cluster 1 ≠ Cluster 2 (p value)	0.00			
Cluster 1 < Cluster 2 (p value)	0.99			
Cluster 1 > Cluster 2 (p value)	0.00			

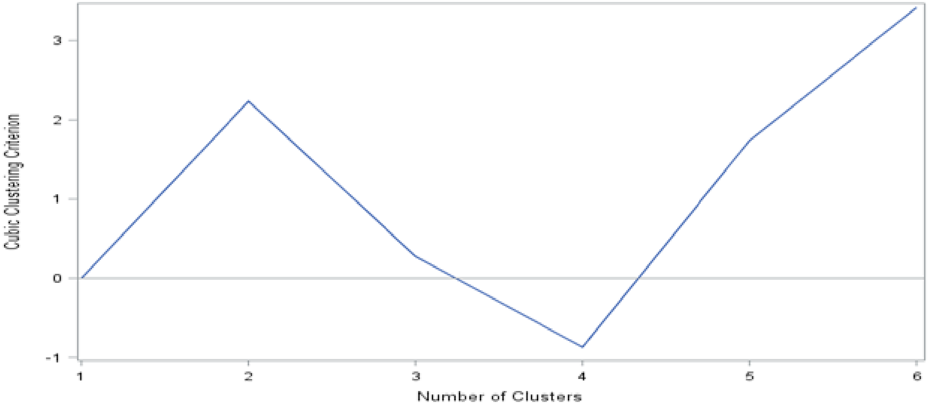
Note: Table 6 exhibits the mean value of cluster 1 and cluster 2. It also gives the t test statistic for the hypothesized difference in the mean value. The p value for the null hypothesis is displayed for both unequal and equal variances. The results of table 6 clearly indicates that stocks of banking firms placed in cluster 1 are statistically different from those placed in cluster 2. Further, the stocks of cluster 1 generate higher returns when compared to the stocks of cluster 2.

Hence, the grouping of stocks on the basis of average distance manifest itself in them being similar their within group elements and dissimilar to their between group elements.

b. K Means Clustering Using Ten Variables

Another method of categorization is the non-hierarchical k-means clustering. It is an exploratory form of categorization used when the researcher is not sure about the number of clusters present in the dataset.

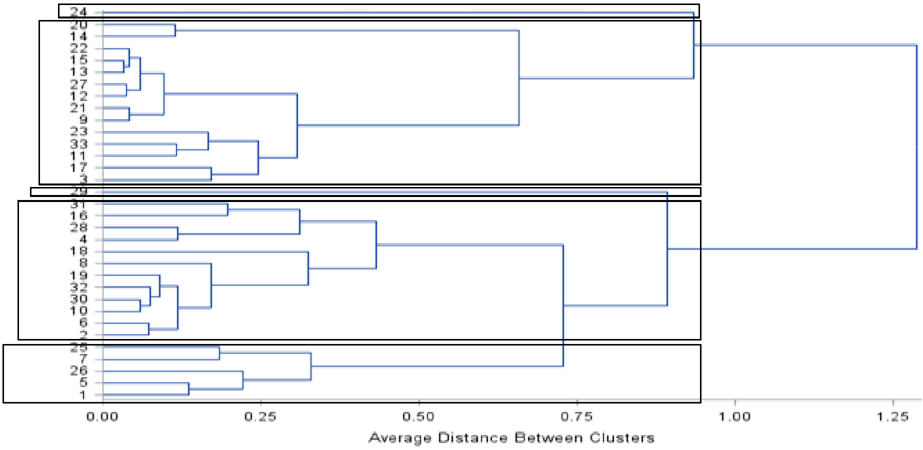
Figure 5: Cluster formation criteria.



Note: Figure 5 shown above the criteria for the number of cluster formation is exhibited. Sharp declining kinks in the value of CCC are used for estimating the number of clusters. In the figure, there is a local peak of the CCC when the number of clusters is five.

It serves as a robustness check mechanism of hierarchical clustering method. The algorithm selects a certain number of clusters randomly and would assign elements of the dataset to their nearest cluster centre using distance measure. Further, the centroid of each cluster is measures and the above mentioned process repeats until a better solution is found. The two dimensional contour of CCC, Pseudo F statistic and Pseudo t^2 is presented graphically in figure 5. Observing the value of CCC from right to left, we can find a sharp decline when number of clusters is 5 as there is a sharp decline in the elbow at cluster number 5.

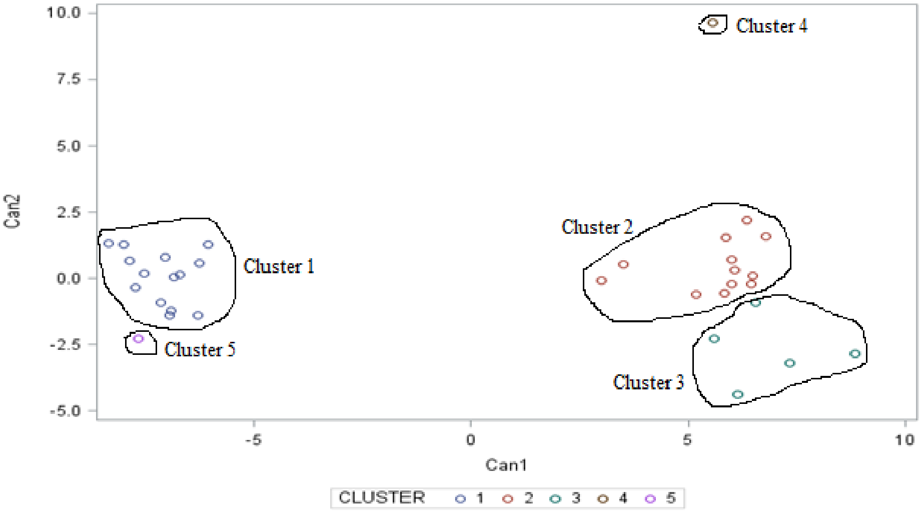
Figure 6: Dendrogram of clusters with average distance between clusters.



Note: Figure 6 exhibits a dendrogram which provides a graphical view of the information in Figure 2. The dendrogram exhibits the average distance between clusters where two data points or sub-clusters join. The figure above clearly exhibits the existence of five distinct clusters.

The Eigen value covariance matrix and cluster history table remain the same as that of hierarchical cluster i.e. Table 3 and 4 respectively. The dendrogram of cluster with average distance between clusters for k-means cluster technique remain the same as that of the hierarchical cluster technique as shown in figure 6. The dendrogram shown in Figure 6 shows the distance at which different stocks are grouped in a cluster. In this clustering technique also, the banks are clustered in five groups on the basis of the variables used in the study.

Figure 7: Clusters for Banking company dataset.



Note: Figure 7 shows the plot of the first two canonical variables (Can1 and Can2) of the five groups formed by average distance measure, considering 33 banking companies used in the study.

In order to analyse the goodness and accuracy of the clusters determined through hierarchical techniques, the variables are plotted on a scatter plot to see their pattern of overlap. As it is not possible for a human brain to visualize and comprehend a ten dimensional vector space, a data reduction technique called as the canonical discriminant analysis is used. In SAS, the can disc procedure is used to generate the variables that creates new variables that are linear combinations of the above ten variables.

As the first two canonical discriminant variables account for most of the variance, can1 and can2 are used as X axis and Y axis of the graph respectively. It is clearly evident from figure 7 that there is little overlap among the five clusters. Hence, the 33 stocks can be clustered into five groups on the basis of the ten variables reduced to canonical discriminant lines. The five clusters for banking companies derived from the hierarchical and k means clustering technique are given in Figure 7.

c. Robustness Check

The robustness of the results found in terms of number of clusters formed can be tested through two sample t test. The clusters formed by the cluster analysis are based on the principle that similar elements are grouped together in a cluster. Hence, we can expect that the mean value of different clusters is different from each other. So, if it can be proved statistically, then it would mean that elements in different clusters are statistically different from each other.

First, we formulate the null hypothesis for a two tailed t test as below:

Null Hypothesis: $\mu_{Cluster\ 1} - \mu_{Cluster\ 2} = 0$

Alternate Hypothesis: $\mu_{Cluster\ 1} - \mu_{Cluster\ 2} \neq 0$

In Table 9 the t coefficients of a two sample t test is shown as robustness measure of clustering. The results of this test indicate that the mean values of stocks belonging to different clusters are different. This establishes that the clustering process has grouped dissimilar companies in different clusters. The results of table 9 clearly indicates that stocks of banking firms placed in cluster 1 are statistically different from those placed in cluster 2.

Table 9: T test table for cluster 1 and 2.

	N	**Mean**	**Std Dev**	**Std Error Mean**
Cluster 1	14	0.003	0.0026	0.001
Cluster 2	17	-0.001	0.0028	0.001
Observed difference (Cluster 1 - Cluster 2)	0.004			
Standard deviation of difference	0.001			
	Unequal Variances			
Degree of freedom	28			
95% Confidence interval for the difference	0.002; 0.006			

t test statistic	4
Cluster 1 ≠ Cluster 2 (p value)	0.00
Cluster 1 < Cluster 2 (p value)	0.99
Cluster 1 > Cluster 2 (p value)	0.00
	Equal Variances
Degree of freedom	29
95% Confidence interval for the difference	0.002; 0.006
t test statistic	4.10
Cluster 1 ≠ Cluster 2 (p value)	0.00
Cluster 1 < Cluster 2 (p value)	0.99
Cluster 1 > Cluster 2 (p value)	0.00

Note: Table 9 exhibits the mean value of cluster 1 and cluster 2. It also gives the t test statistic for the hypothesized difference in the mean value. The p value for the null hypothesis is displayed for both unequal and equal variances. The results of table 10 clearly indicates that stocks of banking firms placed in cluster 1 are statistically different from those placed in cluster 2. Further, the stocks of cluster 1 generate higher returns when compared to the stocks of cluster 2.

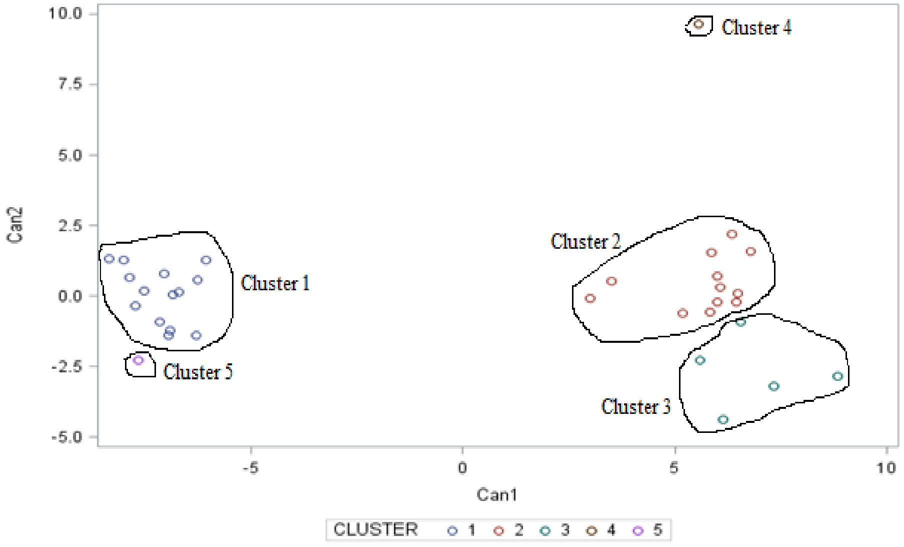
Further, the stocks of cluster 1 generate higher returns when compared to the stocks of cluster 2. Observing the t test statistic from Table 9, it can be asserted that the values of elements in cluster 1 and cluster 2 are unequal. Also, the return generated by stocks of cluster 1 are larger than that of cluster 2.

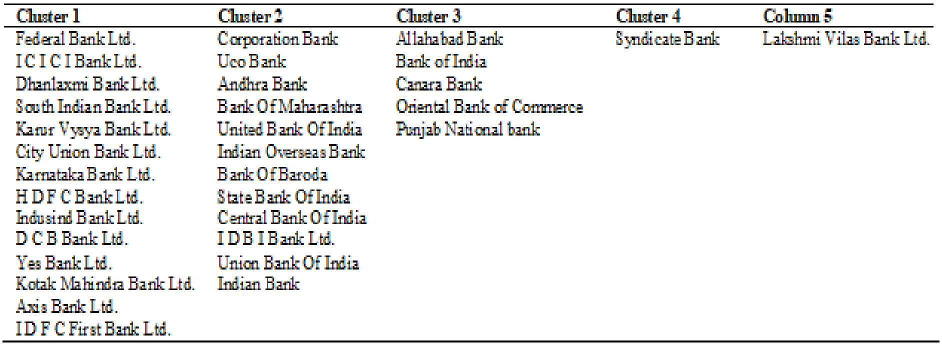
6. Conclusions

The article examines a dataset containing 33 listed banking enterprises and clustered them into five groups on the basis of ten important variables (given in the sub section “variables construction” under the section “Data and Methodology”). The elements within a cluster represent banks homogeneous to each other based on the variables. An investor has the flexibility to choose any of the banks from a cluster as they are all similar to one another. S/he doesn’t have to individually determine the profile of each bank from within a cluster.

Note: Figure 8 shows the plot of the first two canonical variables (Can1 and Can2) of the five groups formed by average distance measure, considering 33 banking companies used in the study. The first cluster contains fourteen companies; the second cluster contains twelve companies; the third cluster contains five companies; the fourth and fifth clusters contain one company each.

Figure 8: Clusters for Banking company dataset along with their names:





Cluster 1	Cluster 2	Cluster 3	Cluster 4	Column 5
Federal Bank Ltd.	Corporation Bank	Allahabad Bank	Syndicate Bank	Lakshmi Vilas Bank Ltd.
I C I C I Bank Ltd.	Uco Bank	Bank of India		
Dhanlaxmi Bank Ltd.	Andhra Bank	Canara Bank		
South Indian Bank Ltd.	Bank Of Maharashtra	Oriental Bank of Commerce		
Karur Vysya Bank Ltd.	United Bank Of India	Punjab National bank		
City Union Bank Ltd.	Indian Overseas Bank			
Karnataka Bank Ltd.	Bank Of Baroda			
H D F C Bank Ltd.	State Bank Of India			
Indusind Bank Ltd.	Central Bank Of India			
D C B Bank Ltd.	I D B I Bank Ltd.			
Yes Bank Ltd.	Union Bank Of India			
Kotak Mahindra Bank Ltd.	Indian Bank			
Axis Bank Ltd.				
I D F C First Bank Ltd.				

An investor has the flexibility to select: either of the bank's stock from the fourteen stocks in cluster 1; either of the bank's stock from the twelve stocks in cluster 2; either of the bank's stock from the five stocks in cluster 3. All the stocks within a cluster have similar characteristics with respect to the variables of interest. Hence, their homogeneity would ensure that inclusion of either one of the stock would have similar impact on the overall portfolio, sans the effort of analysing each company individually.

7. Limitations and Scope for Further Study

One of the biggest limitations of the method is that it is applicable only on similar entities and cannot be applied to different set of units. It can only be applied to homogeneous set of units. There is ambiguity in finding the closest point to the units taken which in real scenario can be overruled. The findings and results of the method are not universal and subject to data set considered. This study is based on the data of one year and only for the banking companies listed in India. This can further be extended to other sectors and multiple years whereby the portfolio manager/investor can fashion a portfolio in consonance with his risk appetite. The stocks that fit into that framework would be considered as a potential stock for addition in the portfolio.

References

Ahamed, N. and Bhattacharjee, K. (2012): Classification of fixed income securities exchange: Clustering and Profiling, *Amity Business Review*, **13 (2)**, 10-18.

Aldenderfer, M. S. and Blashfield, R. K. (1984): Quantitative Applications in the Social Sciences: Cluster analysis. *Thousand Oaks, CA: SAGE Publications, Inc.* doi: 10.4135/9781412983648

Alexandra, H., Joldeș, C. and Dumitrescu, D. (2008) : A Cluster Analysis of Financial Performance in Central and Eastern Europe, *Finances, Banks and Accountancy,* **3**, 301-306

Bensmail, H. and DeGennaro, Ramon P. (2004): Analyzing imputed financial data: a new approach to cluster analysis, *Working Paper: Federal Reserve Bank of Atlanta.*

Das, N. (2003): Hedge Fund Classification Using K-Means Clustering Method, 9th International Conference on Computing in Economics and Finance, *University of Washington, Seattle*

Gibson, Rajna and Sébastien, Gyger (2007): The Style Consistency of Hedge Funds, *European Financial Management,* **13(2)**, 287-308

Jain, A. and Subramanyam, G. (2015): A taxonomy for evaluation and comparison of financial performance of Indian IT companies, *Adarsh Journal of Management Research*, **8(2)**, 1-13.

Jensen, R. E. (1969): A dynamic programming algorithm for cluster analysis. *Operations Research*, **17(6)**, 1034-1057.

Jensen, Robert E. (1971): A Cluster Analysis Study of Financial Performance of Selected Business Firms, *The Accounting Review*, **46(1)**, 36-56

Cormack, R. (1971). A Review of Classification. *Journal of the Royal Statistical Society. Series A (General),* **134(3)**, 321-367.

Kazemi, H., Gupta, B. and Daglioglu, A. (2003): Hedge Fund Classification Methods, Working paper, *Isenberg School of Management, University of Massachusetts, Amherst*

King, Benjamin F. (1966): Market and Industry Factors in Stock Price Behavior, *The Journal of Business*, **39(1)**, 139-190

Martin, George (2001): Making Sense of Hedge Fund Returns: What Matters and What Doesn’t, Derivatives Strategy, *Working paper: Isenberg School of Management, University of Massachusetts, Amherst*

Meyers, Stephen L. (1973): A Re-Examination of Market and Industry Factors in Stock Price Behavior, *Journal of Finance*, **28(3)**, 695-705

Miceli, M. A. and Susinno, G. (2004): Ultrametricity in Fund of Funds Diversification, *Physica A*, **344(1)**, 95-99.

Haldar *et al.* (2008): Cluster Analysis and Clinical Asthma Phenotypes, *American journal of respiratory and clinical care medicine*, **178(3)**, 218-224.

Appendix

List of banks used for the analysis	List of banks deleted from the dataset
Allahabad Bank	Dena Bank
Andhra Bank	Bandhan Bank
Axis Bank	Vijaya Bank
Bank of Baroda	RBL Bank
Bank of India	Punjab & Sind Bank
Bank of Maharashtra	Jammu & Kashmir Bank
Canara Bank	AU Small Finance Bank
Central Bank of India	
City Union Bank	
Corporation Bank	
DCB Bank	
Dhanlaxmi Bank	
Federal Bank	
HDFC Bank	
ICICI Bank	
IDBI Bank	
IDFC Bank	
Indian Bank	
Indian Overseas Bank	

IndusInd Bank Karnataka Bank Karur Vysya Bank Kotak Mahindra Bank Lakshmi Vilas Bank Oriental Bank of Commerce Punjab National Bank South Indian Bank State Bank of India Syndicate Bank UCO Bank Union Bank of India United Bank of India Yes Bank	

Note: The table above exhibits the list of banks selected and rejected for our analysis.

Table 10: t test table for cluster 1 and 3.

	N	Mean	Std Dev	Std Error Mean
Cluster 1	14	0.003	0.0026	0.001
Cluster 3	5	0.001	0.0017	0.001
Observed difference (Cluster 1 - Cluster 3)	0.002			
Standard deviation of difference	0.001			

Unequal Variances	
Degree of freedom	11
95% Confidence interval for the difference	-0.0002; 0.0042
t test statistic	2
Cluster 1 ≠ Cluster 3 (p value)	0.07
Cluster 1 < Cluster 3 (p value)	0.96
Cluster 1 > Cluster 3 (p value)	0.03
Equal Variances	
Degree of freedom	17
95% Confidence interval for the difference	-0.0001; 0.0041
t test statistic	1.59
Cluster 1 ≠ Cluster 3 (p value)	0.12
Cluster 1 < Cluster 3 (p value)	0.93
Cluster 1 > Cluster 3 (p value)	0.06

Note: Table 10 exhibits the mean value of cluster 1 and cluster 3. It also gives the t test statistic for the hypothesized difference in the mean value. The p value for the null hypothesis is displayed for both unequal and equal variances. The results of table 10 clearly indicates that stocks of banking firms placed in cluster 1 are statistically different from those placed in cluster 3. Further, the stocks of cluster 1 generate higher returns when compared to the stocks of cluster 3.

The results of table 10 clearly indicates that stocks of banking firms placed in cluster 1 are statistically different from those placed in cluster 3. Further, the stocks of cluster 1 generate higher returns when compared to the stocks of cluster 3.

Table 11: t test table for cluster 2 and 3.

	N	**Mean**	**Std Dev**	**Std Error Mean**
Cluster 2	12	-0.002	0.0028	0.001
Cluster 3	5	0.001	0.0017	0.001
Observed difference (Cluster 2 - Cluster 3)	-0.003			
Standard deviation of difference	0.0011			
Unequal Variances				
Degree of freedom	12			
95% Confidence interval for the difference	-0.0054; -0.0006			
t test statistic	-2.72			
Cluster 2 ≠ Cluster 3 (p value)	0.01			
Cluster 2< Cluster 3 (p value)	0.99			
Cluster 2> Cluster 3 (p value)	0.00			
Equal Variances				
Degree of freedom	15			
95% Confidence interval for the difference	-0.0053; -0.0007			
t test statistic	-2.16			
Cluster 2 ≠ Cluster 3 (p value)	0.04			
Cluster 2< Cluster 3 (p value)	0.97			
Cluster 2> Cluster 3 (p value)	0.02			

Note: Table 11 exhibits the mean value of cluster 2 and cluster 3. It also gives the t test statistic for the hypothesized difference in the mean value. The p value for the null hypothesis is displayed for both unequal and equal variances. The results of table 11 clearly indicate that stocks of banking firms placed in cluster 2 are statistically different from those placed in cluster 3. Further, the stocks of cluster 2 generate higher returns when compared to the stocks of cluster 3.

The results of table 11 clearly indicate that stocks of banking firms placed in cluster 2 are statistically different from those placed in cluster 3. Further, the stocks of cluster 2 generate higher returns when compared to the stocks of cluster 3.

The t test results for other clusters using different measures of distance are not shown here in the interest of parsimony. The mean return for elements in a cluster is statistically different than the elements in another cluster.

Table 1: Standardized data.

Variable	**Mean**	**Std. Dev**	**Maxi.**	**Min.**	**25th Pctl**	**50th Pctl**	**75th Pctl**
Age	77.91	38.34	154	5	34	91	108
Prom	45.77	36.08	92.25	0	8.88	52.02	81.73
CR	4.62	2.13	9.48	0.93	3.02	4.28	6.28
D/E	1.31	1.02	4.03	0.21	0.6	0.95	1.67
PAT	-7779.07	53861.21	210781.7	-99754.9	-37378.8	549.9	5920
TA	4466797.3	6523848.7	36809142	122893.6	1596531.8	2500084	4965523
TL	4466797.3	6523848.7	36809142	122893.6	1596531.8	2500084	4965523
Cash flow	-22548.25	188967.83	384187.9	-560546.6	-108226.9	-22310.5	29099
Deposit	3517718.7	5121505.6	29113860	109196.6	1349543.4	2225341	4159153
GOI	45.28	42.73	96.83	0	0	63.26	87.05
Revenue	367150.1	510488.07	2834413.2	11205.1	127700.6	219865.2	402841.6
Beta	1.6	0.45	2.38	0.74	1.34	1.65	1.98
M_Cap	580516.41	1179695.6	5769966.1	4250.6	60524.21	108942.6	314287.3
EPS	-16.03	39.28	69.48	-101.01	-31.74	-10.28	4.77
P/B	1.4	1.24	5.85	0.43	0.67	0.82	1.45
ROA	0.09	0.01	0.1	0.07	0.08	0.09	0.09
Return	0	0	0.01	-0.01	0	0	0
Risk	0.05	0.01	0.08	0.02	0.04	0.05	0.06

Note: Table 1 exhibits the descriptive statistics of variables. The first column indicates the mean value, second column indicates standard deviation, third column indicates maximum value, fourth column indicates minimum value, fifth column indicates 25th percentile, sixth column indicates 50th percentile or median and seventh column indicates the 75th percentile values. Age refers to the age of the firm is calculated by subtracting the incorporation year from current year. Age refers to the age of the firm is calculated by subtracting the incorporation year from current year. Prom refers to the percentage of promoters holding in the firm and is calculated by dividing the promoters holding by total holding of the firm. CR refers to current ratio and is calculated by dividing the current assets by current liabilities. D/E refers to the debt to equity ratio of the firm is calculated by long term debt by equity capital. PAT refers to the profit after tax of the firm. TA refers to the total assets of the firm. TL refers to the total liabilities of the firm. Cash flow refers to the cash flow from operating activities of the firm. Deposit refers to the total amount accepted by the bank in the form of deposits for the current year. GOI refers to the percentage of holding by the government of India in the bank. Revenue refers to the amount of revenue generated by the bank. Beta refers to the sensitivity of the stock with respect to that of the market. M_Cap refers to the market capitalization which is calculation as the product of the market price of the stock with the total number of outstanding shares. EPS refers to the earnings per share and is calculated by dividing the profit after tax by the number of shares outstanding. P/B refers to the price to book ratio, Return refers to the return generated by the stock and risk refers to the risk embedded in the stock i.e. the squared deviation from the expected mean value.

Table 2: Standardized data.

Comp	Prom (%)	CR	D/E	GOI	Beta	EPS	P/B	ROA	Return	Risk
Allahabad Bank	79.41	4.12	1	85.8	2.03	-61.49	1	0.0814	-0.0006	0.0609
Andhra Bank	84.83	5.4	0.8	90.9	1.98	-23.6	0.5	0.0849	-0.0029	0.056
Axis Bank Ltd.	23.71	3.02	2.3	0	1.34	-0.38	2.4	0.0851	0.0038	0.0387
Bank Of Baroda	63.74	6.73	1.5	63.3	1.67	-10.08	0.8	0.0721	-0.0005	0.0588
Bank Of India	83.09	8.22	1.1	87.1	2.14	-73.49	0.8	0.075	0.00263	0.0724
Bank Of Maharashtra	87.01	2.51	2.3	87.7	1.36	-18.83	1	0.0774	-0.0046	0.0502
Canara Bank	72.55	6.28	1.4	70.6	2.1	-99.06	0.7	0.0812	0.00295	0.0644
Central Bank Of India	88.02	7.19	0.3	91.2	1.47	-31.65	0.8	0.0854	-0.0055	0.0705
City Union Bank Ltd.	0	3.36	0.4	0	0.78	9.01	3.1	0.0985	0.0064	0.0366
Corporation Bank	86.77	2.13	0.5	93.5	1.45	-16.49	0.6	0.0909	-0.0022	0.0509
D C B Bank Ltd.	14.93	2.65	1	0	1.23	9.33	1.9	0.0947	0.0054	0.0411
Dhanlaxmi Bank Ltd.	0	6.97	0.5	0	1.78	-2.2	0.6	0.0912	-0.0005	0.0776
Federal Bank Ltd.	0	6.22	0.6	0	1.42	4.35	1.4	0.08	0.00423	0.043
H D F C Bank Ltd.	26.54	2.43	0.8	0	0.74	69.48	4.4	0.0937	0.00557	0.0191
I C I C I Bank Ltd.	0	3.42	1.6	0	1.48	4.77	2.2	0.0805	0.00512	0.0401
I D B I Bank Ltd.	52.02	4.21	3.9	81	2.01	-31.74	1.4	0.0904	0.0005	0.0589
I D F C First Bank Ltd.	40	0.93	3.9	0	1.49	-1.86	1.4	0.0783	-0.0015	0.0491

Table 2 to continue.

Indian Bank	81.73	5.85	0.7	81.5	2.03	8.62	0.7	0.0775	0.00323	0.0645
Indian Overseas Bank	91.99	9.48	0.4	92.5	1.56	-10.28	0.8	0.0879	-0.004	0.0402
Indusind Bank Ltd.	16.79	3.03	1.6	0	1.14	64.54	3.6	0.0993	0.00356	0.027
Karnataka Bank Ltd.	0	4.31	0.6	0	1.82	11.96	0.6	0.0874	0.00119	0.0432
KarurVysya Bank Ltd.	2.11	3.65	0.2	0	0.97	1.89	1.1	0.0981	0.00317	0.0381
Kotak Mahindra Bank Ltd.	30.01	1.87	0.8	0	0.9	21.15	5.9	0.0916	0.00563	0.0274
Lakshmi Vilas Bank Ltd.	8.88	3.2	1.9	0	0.95	-71.19	1.5	0.0974	-0.0028	0.0494
Oriental Bank Of Commerce	77.23	5.23	0.8	87.6	2.31	-101.01	0.6	0.0891	0.0011	0.0601
Punjab National Bank	70.22	5.88	1	75.4	1.95	-80.73	0.7	0.0835	-0.0007	0.0766
South Indian Bank Ltd.	0	6.88	1	0	1.12	1.4	0.5	0.0825	-5.00E-05	0.0513
State Bank Of India	58.53	4.28	2.1	57.1	1.65	-17.93	1.4	0.077	0.00268	0.0464
Syndicate Bank	0	1.95	1.7	84.7	2.2	-31.8	0.6	0.0792	-0.0021	0.0611
Uco Bank	90.8	4.66	0.6	93.3	1.95	-13.44	0.8	0.0694	-0.003	0.0481
Union Bank Of India	67.43	6.95	1.8	74.3	2.38	-40.13	0.4	0.0811	-0.0004	0.0704
United Bank Of India	92.25	7.21	0.2	96.8	1.65	-17.25	0.7	0.0895	-0.0036	0.0472
Yes Bank Ltd.	19.82	2.28	4	0	1.66	19.06	1.5	0.0898	0.00054	0.0678

Note: Table 2 above exhibits the standardized value of variables. The standardization is done to mould every variable with same variance. A standardized value undergoes a form of scaling or transformation such that those values would have a mean of zero and standard deviation of one.