IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN:2319-765X. Volume 10, Issue 1 Ver. IV. (Feb. 2014), PP 63-71
www.iosrjournals.org

Visualization Map and Evaluation of Indian Industries Using Statistical Data Mining Techniques

R. Chandrasekaran (Retd.)[1], R.Lakshmi Priya[2] and G.Manimannan[3]
[1 and 3]*Department of Statistics Madras Christian College, Chennai-59.*
[2] *Department of Statistics Dr. Ambedkar Government Arts College, Chennai*

Abstract: *The uncovering of company and financial difficulties is a theme which has particularly susceptible to financial ratio analysis. Rating and predictive the show of the top ranking companies on the basis of certain financial ratios based on Statistical Data Mining (SDM) techniques. In this paper we used three different Statistical Data Mining methods and they are Factor Analysis, K-means Clustering techniques and Multivariate Disriminant Analaysis. It is well known that statistical information on financial ratios is being extensively used by researchers for many purposes. The financial information of public and private sector companies rated as the best with reference to net sales, it was published by capital market, were considered for the period from 2001 to 2010 for the present study. Out of many possible ratios, eight financial ratios with different notions of the objectives and significant meaning in the literature were selected. To explore the financial data, data mining tools such as factor analysis, k-means clustering and discriminant analyses are applied in succession. Factor analysis is initiated first to uncover the structural models underlying financial ratios. The scores from extracted factors were then used to find initial groups by k-means clustering algorithm to prune the data. The cluster analysis was followed by iterative discriminant procedure with original ratios until cent percent classification was achieved. Finally, the groups were identified as companies belonging to Grade H, Grade M and Grade L in that order, which show the behavior of High performance, Moderate performance and Low performance. Similarly the companies belonging to Grade M category are superior to those of Grade L but inferior to Grade H, indicating the members in the category Grade L are at the low profile in terms of the importance of some of the parameters considered for the present study. The pictorial representations of the groups are also depicted. Industrial analyst can make use of the grading of industries proposed in this paper to know the performance of companies and for the purpose of investment.*

Key Words: *Financial Ratios, Factor Analysis, k-mean Clustering, Discriminant, Analysis, Industrial Performance.*

I. Introduction

Corporate failures are a common problem of developing and developed economies (Atlman *et.al*., 1979). Now Days Industry is the segment of economy concerned with the production of goods and services. Industry began in its present form during the 1800s, aided by technological advances, and has continued to develop into new types and sectors. An industrial growth is phased on three stages; they are birth, growth, and maturity. Its management should attempt to prolong the growth stage through acquisitions, research and development of new products. But unfortunately, some industries, even if they use strategies, fail each year. A business failure is an unfortunate event. There are a variety of cases where majority of industries may fail within the first or second year of its life and other industries may grow, mature and perform much later. Industries may fail, if returns are too low or if they become technically insolvent or if they become bankrupt. Bankruptcy is an oblivious form of failure. It occurs when a firm's liabilities exceeds the fair market value of its assets. The other major causes of business failure can be attributed to lack of managerial skill, economic activity and corporate maturity. At this stage of failure, the firm becomes answerable to its creditors and shareholders. So, a good management becomes more responsible to help the firm in avoiding ultimate failure. This implies that they should be capable of predicting the firm's performance as much early as possible.

II. Review Of Literature

Financial analysis is the utilization of a group of analytical techniques, to determine the strengths, weaknesses and direction of a company's performance. Financial ratio is a statistical yardstick that relates two numbers generally taken from a firm's income statement, balance sheet or both at a specific point in time. Selection of the ratios depends upon the purpose of the analysis. A carefully executed financial statement analysis can assist the financial manager in assessing financial conditions of the firm. Financial ratios have been used as inputs for advanced statistical models to forecast many kinds of business events and to identify financial and other characteristics, but the focus has been on testing statistical models to predict business failure. The pioneering work initiated in this area of research was by Altman (1983). Chandrasekaran *et. al.* (2013), taking

the financial ratios of the top ranking companies in India as the variables, using data mining techniques such as self organizing principle and discriminate analysis, companies performance perception were analyzed and found the best performed companies. Chandransekaran *et. al.* (2013) rated the top ranking companies based on financial parameters. The present research aims at analyzing financial information of public and private sector companies, rated as the best with reference to net sales during the period 2001 to 2010.

III. Database And Variables

The data was collected from capital market in Chennai. The data mainly consists of the five major industrial sectors of Cement, Cotton, Hotel, Leather and Sugar industries in India. Under each category there are several companies, and about 890 companies where taken for the present study. The data had financial ratios of each company for the time period of ten years (from 2001 to 2010). Initially, fourteen parameters are considered for this study, and some of them are discarded that did not vary over in the time period considered for study. Finally, eight financial ratios are obtained. The eight ratios are (1) Debt-Equity Ratio (DER), (2) Long Term Debt-Equity Ratio (LTDER), (3) Current Ratio (CURR), (4) Fixed Assets (FIXA), (5) Inventory (INV), (6) Debtors Ratio(DEB), (7) Return on Capital Employed (ROCE), (8) Return on net worth (RONW). We have normalized the parameters using total output. The objectives of the present study are: (i) to determine the hidden patterns or classes of industries based on financial ratios using factor analysis. (ii) to propose different methods of grade the industries on the basis financial ratios, using k-mean clustering technique. (iii) to cross validate the previous method Multivariate Discriminant Analysis tool, based on the underlying classification maps.

IV. Result And Discussion

IV. i. Factor Analysis

Factor analysis is a statistical data reduction technique used to explain variability among observed random variables in terms of fewer unobserved random variables called factors. The observed variables are modeled as linear combinations of the factors, plus "error" terms (Johnson and Wichern, 1992). The main objectives of factor analysis are to reduce the number of variables and to detect structure in the relationships between variables. Therefore, factor analysis is applied as a data reduction or structure identification method. Although there are several techniques of data reduction, factor analysis is by far the most frequently used method in financial researches (Mahmoud, Judith and Cecilio, 1987).

IV. ii. Principal Component Analysis

Principal Component Analysis (the total variance in the data is considered) and Common factor analysis (the common variance is considered) are the two approaches of factor analysis. In the present study, we apply Principal Component Analysis (PCA) to verify whether it is possible to reduce the 8 different ratios into a few significant artificial variables. For this purpose we used financial ratios of companies within five major Indian industries. The application of PCA is also extended to verify whether the discrimination of the firms with specific characteristics will pose any problem and also to find the determinant ratios related to each of the principal components.

The proportion of sample variance is prefixed based on the subject matter of interest. The application of PCA revealed consistently three factors for each year with eigen values little less than or equal to unity. Further support for retaining three factors is assessed on studying the residual matrix. All the entries in this matrix are comparatively small during the study periods for the extracted number of factors. Thus we pursued with 3 factor model for the present study. Table 1 show the eigen values and percentage of variances accounted for by each factor and the total variability during the study period from 2001 to 2010. From the below table, we observe that the total variances explained by extracted factors are over 70 percent, which is relatively high. Also, for each factor the variability is more or less the same for the study period. Hence, principal factor extractions with variations of three factors were therefore requested.

TABLE 1. EIGEN VALUES AND PERCENTAGE OF VARIANCES EXPLAINED BY FACTOR ANALYSIS

FACTOR	1998		1999		2000		2001		2002	
	EIGEN VALUE	%VAR	EIGEN VALUE	%VAR	EIGEN VALUE	%VAR	EIGEN VALUE	%VAR	EIGEN VALUE	%VAR
1	2.547	31.834	2.980	37.251	2.832	35.395	2.703	33.789	3.079	38.491
2	1.831	22.885	1.672	20.899	1.812	22.648	2.023	25.285	1.718	21.473
3	1.307	16.337	1.082	13.522	1.196	14.946	1.202	15.030	.920	11.498
TOTAL	71		72		73		74		72	

FACTOR	2003		2004		2005		2006		2007	
	EIGEN VALUE	%VAR	EIGEN VALUE	%VAR	EIGEN VALUE	%VAR	EIGEN VALUE	%VAR	EIGEN VALUE	%VAR
1	3.062	38.280	2.941	36.763	2.308	28.845	2.286	28.576	2.698	33.730
2	1.546	19.327	1.568	19.598	1.903	23.783	1.914	23.927	1.663	20.787
3	1.068	13.353	1.135	14.182	1.419	17.737	1.413	17.662	1.329	16.618
TOTAL	71		71		70		70		71	

In fact, in studying the characteristics of the companies based on eight financial parameters, the application of PCA facilitated characterizing the properties of the top ranking industries in the form of reduced structure called factor scores. These extracted scores were then subjected to further investigations. In the unrotated factor loading matrix, certain ratios are highly correlated with one factor and to a considerable extent with other factors.

IV. iii. Rotation of Factor Scores

The factor loading accompanying rotation method is not always easy to interpret. Rotation of factor is subjected until a simplest structure is achieved. Therefore rotations are tried with more than one method. In all that varimax rotation provided relatively better pattern of financial ratios and is incorporated in the present study. The specific goal of varimax rotation is to make the factors as simple as possible by maximizing the variance of the loading across the variables within factors. In this way loading tend to become higher for those variables with high correlation with the factor and smaller for other variables. After identifying the significant factor loadings an attempt is made to assign some meanings to the factors based on the patterns of the loadings. In general, the larger the absolute size of the factor loading for a variable, the more important is the variable in interpreting the factors. At the same time due importance are given to those variables with sign loading in a factor for naming them intelligently. Hence, factor loading or factor structure matrix is a m by n matrix of correlation between the original variable and their factors. On the whole, the clustering of financial ratios in 3 factors extracted with varimax rotation is summarized in following Table 2.

Table 2. Summary of variable loading in Rotated Factor (Year wise)

No	Factor Name	2001	2002	2003	2004	2005
1	**Debtors**	DER, LTDER, CURR, DEB	ROCE, RONW, FA	DER, LTDER, DEB, CURR	DER, LTDER, DEB, CURR	ROCE, FA, RONW,
2	**Profitability**	ROCE, RONW	DER, LTDER, CURR, DEB	ROCE, RONW	ROCE, RONW	DER, LTDER, INV

3	**Asset / Liquidity**	INV, FA	INV	INV, FA	FA, INV	DEB, CURR
No	**Factor Name**	**2006**	**2007**	**2008**	**2009**	**2010**
1	**Debtors**	DER, LTDER, DEB	DER, LTDER, DEB	LTDER, DER	DER, LTDER, DEB	RONW, ROCE, INV
2	**Profitability**	RONW, ROCE	ROCE, RONW	ROCE, RONW	ROCE, RONW, FA	LTDER, DER
3	**Asset / Liquidity**	INV, CURR, FA,	INV, CURR, FA,	CURR, FA, INV, DEB	INV, CURR	CURR, FA, DEB

IV. iv. Results of Factor Analysis

The naming of factors comes from the meanings associated with the set of variables grouped together for each factor. In the present study, factor loading of variables clustered are ascended from the top with the largest loading and so on in each column against their respective factor. Thus, financial ratios highly correlated with each factor can be observed from Table 2. On the other hand, we observed slight changes in clustering of variables on factors during the study periods. This difference in the set of variables in factor may be due to financial strength of the country over the year. Table 2 also gives some indication on the extent of stability in financial patterns over the years from 2001 to 2010, in terms of the consistency of loadings on factors. Overall, loadings on the extracted three factors appear to be stable during the study period even though slight changes in the clustering are reported. Mainly, the change in variables clustered or loading may be due to statistical variations in the original data although the underlying structure remained the same. Yet, we also found that the loading of certain ratios during the study periods declined slowly.

These eight variable are grouped into three factors on the basis of the inter relationship among themselves. The ratios like *Debt-equity ratio*, *Debtors* and *Long term debt-equity Ratio* are highly loaded in the first factor and hence they form a group. For the year 2002, 2004, and 2010 these ratios are loaded in the second factor. As these show that the ratios are profitability of the industries, this factor is named as *debtor's factor*. It is interesting to note that this factor contributes from 28 to 38 percent of the total variation. The ratios *Return on capital employed* and *Return on net worth* are found to play a vital role in second factor. These ratios are mainly used in examining the profit of an industry. For the year 2002, 2004 and 2010, shifts in these ratios were observed which are due to statistical variability in data. After a careful examination, factor 2 is labeled as *profitability factor*. Higher profitability ratios would mean that the performance of the industry is good and intern might attract investors or share on these industries. This factor accounts from 19 to 25 percent of the total variation.

The ratios like *Inventory*, *Fixed asset* and *Current ratio* are loaded highly in third factor. We observed that the ratios Fixed asset and Current ratio are shifted mostly. The ratio *Inventory* is stable in third factor expect the year 2003 and 2007. These ratios measure the efficient utilization of owner's capital and growth towards a company. Therefore, this factor is labeled as *Asset* or *Liquidity* factor. From Table 2, we observed that the stability of the company is more during 2001 and 2010 though their *fixed asset* and *current ratio* utilization are not satisfactory. Higher these ratios mean high return and high future growth capacity of a firm. The third factor accounts from 11 to 17 percent of the total variation. In general, the extracted three factors signify the stability of the companies to certain extent owing to statistical variability in the data without ambiguity.

IV. v. k-mean Clustering and Grading of Industries

This form of clustering is also called rapid clustering or k-mans clustering. It is iterative and processes the solution till we arrive at answers for cluster members and cluster centres. This type of clustering needs an initial hypothesis on the number of clusters. Non-hierarchical clustering needs the following: (i) Number of

clusters and (ii) clustering method. In this study, we use the running means to get the best solutions. A good non-hierarchical clustering will converge in less than 10 iterations. There are two important outputs used from this clustering procedure. They are, (i) Cluster Membership and (ii) Cluster Centers. Thus cluster membership updated us about the number of elements in each cluster. This will inform us about the relative importance of each cluster. Larger clusters indicate greater sections of the industries. The cluster centres describe each group's characteristic which can be used for distinguishing feature between the groups. They can help in the industrial performance of the companies, arranging financials for special ratios, performance years, etc.

In this study, to see the performances of the industries, an attempt is made to classify the industries into different clusters. We used k-means clustering procedure using $2,3,\ldots,n$ clusters. Finally we found that only 3 clusters are meaningful during each of the study period. This facilitated to group the companies into 3 clusters and also to grade those as Grade H (High), Grade M (Medium) and Grade L (Low) based on the group means of certain industrial parameters. Grade H industries are those with higher profile than M and L. Similarly, Grade M industries have their profile of performances superior to Grade L but inferior to Grade H, and Grade L are at a low profile in terms of the industrial parameters considered in the present analysis. The grades are assigned based on the magnitude of mean vectors of clusters for each year taking into account certain industrial financial ratios. These three groups paved the way to rate the members in the clusters as Grade H, Grade M and Grade L.

TABLE 3. CLUSTER MEMBERS BASED ON K-MEAN CLUSTERING

Year	Grade H	Grade M	Grade L
2001	32	12	37
2002	22	25	29
2003	58	18	04
2004	58	05	14
2005	20	47	11
2006	15	21	42
2007	34	06	39
2008	36	37	08
2009	10	18	50
2010	32	15	31

The movements of the industries are investigated in addition to grading the industries. This involves identifying the number of industries that had a movement from one profile to another in the consecutive year. That is, an industry could move from Grade L to Grade M, Grade M to Grade H, or Grade L to Grade H from one year to the next. This would indicate that the movements are in the positive direction and the industries do better in the successive years. From Table 3, we observe that for the years 2001, 2003, 2004, 2003, 2008 and 2010 movement is in positive direction and for the rest of the years, 2002, 2005, 2006 and 2009 there is a negative movements.

V. Multivariate Discriminant Analysis

In the present study, Multivariate Discriminant Analysis is used to exhibit groups graphically and judge the nature of overall performance of the companies. This process re-allocated the companies that were assigned a group label by k-means clusters as a seed point.

Algorithm:

A brief algorithm to grade the industries during each of the study period based on their overall performance is described below:

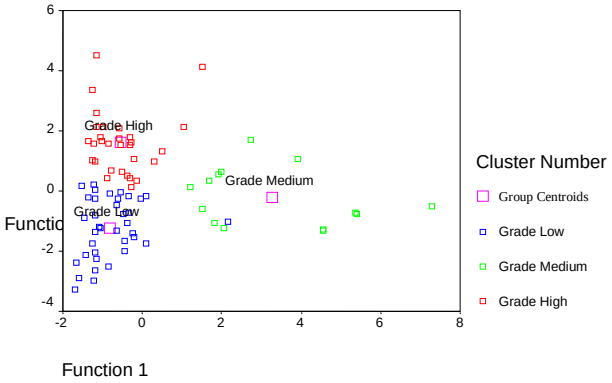
Step 1: k –means analysis partitioned the data set into k-groups using the 3 factor scores as input matrix.

Step 2: Multivariate Discriminant Analysis is then performed with the original industrial parameters by considering the groups formed by the k-means algorithm.

TABLES AND FIGURES 4. PERCENTAGE OF CLASSIFICATION (2001-2010)

Canonical Discriminant Functions

2001

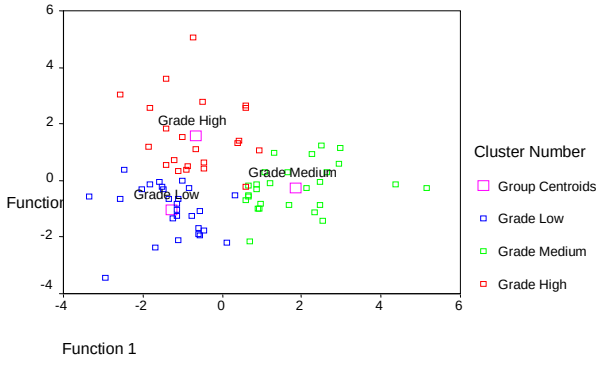


Classification Results

			Predicted Group Membership			
		Cluster Number of Ca	Grade High	Grade Medium	Grade Low	Total
Original	Count	Grade High	29	0	1	30
		Grade Medium	0	14	0	14
		Grade Low	0	1	36	37
	%	Grade High	96.7	.0	3.3	100.0
		Grade Medium	.0	100.0	.0	100.0
		Grade Low	.0	2.7	97.3	100.0
Cross-validated	Count	Grade High	28	1	1	30
		Grade Medium	1	12	1	14
		Grade Low	2	1	34	37
	%	Grade High	93.3	3.3	3.3	100.0
		Grade Medium	7.1	85.7	7.1	100.0
		Grade Low	5.4	2.7	91.9	100.0

Canonical Discriminant Functions

2002

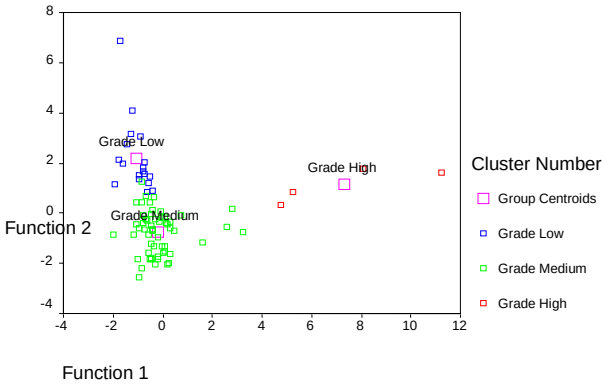


Classification Results

			Predicted Group Membership			
		Cluster Number of Ca	Grade High	Grade Medium	Grade Low	Total
Original	Count	Grade High	20	2	0	22
		Grade Medium	0	27	0	27
		Grade Low	0	1	26	27
	%	Grade High	90.9	9.1	.0	100.0
		Grade Medium	.0	100.0	.0	100.0
		Grade Low	.0	3.7	96.3	100.0
Cross-validated	Count	Grade High	18	2	2	22
		Grade Medium	0	26	1	27
		Grade Low	1	1	25	27
	%	Grade High	81.8	9.1	9.1	100.0
		Grade Medium	.0	96.3	3.7	100.0
		Grade Low	3.7	3.7	92.6	100.0

Canonical Discriminant Functions

2003

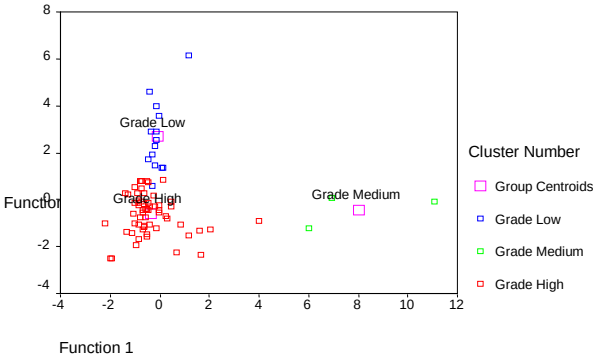


Classification Results

			Predicted Group Membership			
		Cluster Number of Ca	Grade High	Grade Medium	Grade Low	Total
Original	Count	Grade High	4	0	0	4
		Grade Medium	0	57	1	58
		Grade Low	0	0	18	18
	%	Grade High	100.0	.0	.0	100.0
		Grade Medium	.0	98.3	1.7	100.0
		Grade Low	.0	.0	100.0	100.0
Cross-validated	Count	Grade High	3	1	0	4
		Grade Medium	0	56	2	58
		Grade Low	0	2	16	18
	%	Grade High	75.0	25.0	.0	100.0
		Grade Medium	.0	96.6	3.4	100.0
		Grade Low	.0	11.1	88.9	100.0

Canonical Discriminant Functions

2004

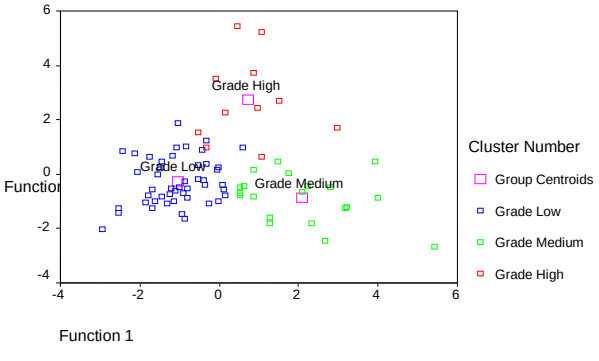


Classification Results

			Predicted Group Membership			
		Cluster Number of Case	Grade High	Grade Medium	Grade Low	Total
Original	Count	Grade High	59	1	0	60
		Grade Medium	0	3	0	3
		Grade Low	1	0	13	14
	%	Grade High	98.3	1.7	.0	100.0
		Grade Medium	.0	100.0	.0	100.0
		Grade Low	7.1	.0	92.9	100.0
Cross-validated	Count	Grade High	59	1	0	60
		Grade Medium	0	3	0	3
		Grade Low	3	0	11	14
	%	Grade High	98.3	1.7	.0	100.0
		Grade Medium	.0	100.0	.0	100.0
		Grade Low	21.4	.0	78.6	100.0

Canonical Discriminant Functions

2005

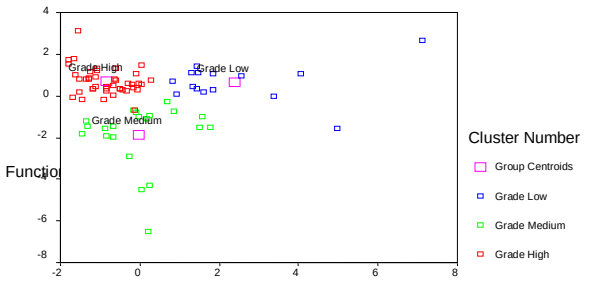


Classification Results

			Predicted Group Membership			
		Cluster Number of Case	Grade High	Grade Medium	Grade Low	Total
Original	Count	Grade High	9	1	1	11
		Grade Medium	0	20	0	20
		Grade Low	2	0	45	47
	%	Grade High	81.8	9.1	9.1	100.0
		Grade Medium	.0	100.0	.0	100.0
		Grade Low	4.3	.0	95.7	100.0
Cross-validated	Count	Grade High	7	2	2	11
		Grade Medium	0	16	4	20
		Grade Low	3	0	44	47
	%	Grade High	63.6	18.2	18.2	100.0
		Grade Medium	.0	80.0	20.0	100.0
		Grade Low	6.4	.0	93.6	100.0

Canonical Discriminant Functions

2006

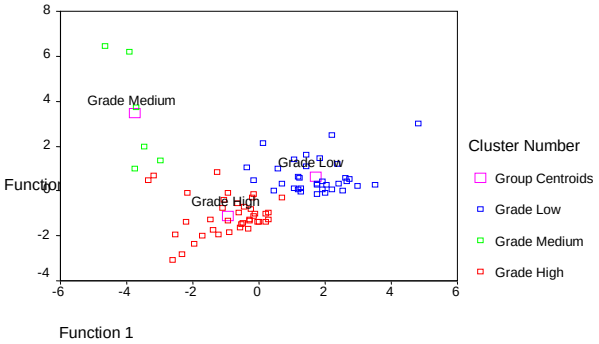


Classification Results

			Predicted Group Membership			
		Cluster Number of Case	Grade High	Grade Medium	Grade Low	Total
Original	Count	Grade High	41	1	0	42
		Grade Medium	0	21	0	21
		Grade Low	0	0	15	15
	%	Grade High	97.6	2.4	.0	100.0
		Grade Medium	.0	100.0	.0	100.0
		Grade Low	.0	.0	100.0	100.0
Cross-validated	Count	Grade High	41	1	0	42
		Grade Medium	2	17	2	21
		Grade Low	2	0	13	15
	%	Grade High	97.6	2.4	.0	100.0
		Grade Medium	9.5	81.0	9.5	100.0
		Grade Low	13.3	.0	86.7	100.0

Canonical Discriminant Functions

2007

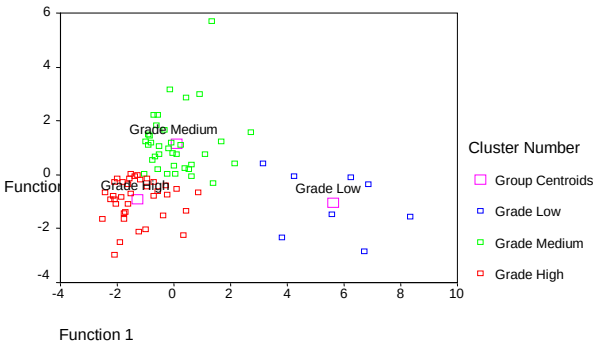


Classification Results[a]

			Predicted Group Membership			
		Cluster Number of Ca	Grade High	Grade Medium	Grade Low	Total
Original	Count	Grade High	37	1	1	39
		Grade Medium	0	6	0	6
		Grade Low	1	0	33	34
	%	Grade High	94.9	2.6	2.6	100.0
		Grade Medium	.0	100.0	.0	100.0
		Grade Low	2.9	.0	97.1	100.0
Cross-validated[b]	Count	Grade High	36	2	1	39
		Grade Medium	2	4	0	6
		Grade Low	2	0	32	34
	%	Grade High	92.3	5.1	2.6	100.0
		Grade Medium	33.3	66.7	.0	100.0
		Grade Low	5.9	.0	94.1	100.0

Canonical Discriminant Functions

2008

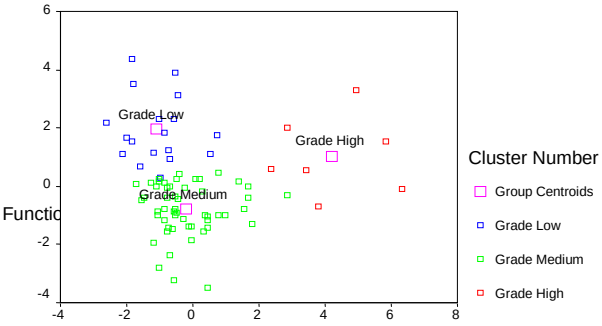


Classification Results[a]

			Predicted Group Membership			
		Cluster Number of Ca	Grade High	Grade Medium	Grade Low	Total
Original	Count	Grade High	36	1	0	37
		Grade Medium	1	35	0	36
		Grade Low	0	0	8	8
	%	Grade High	97.3	2.7	.0	100.0
		Grade Medium	2.8	97.2	.0	100.0
		Grade Low	.0	.0	100.0	100.0
Cross-validated[b]	Count	Grade High	36	1	0	37
		Grade Medium	2	34	0	36
		Grade Low	0	1	7	8
	%	Grade High	97.3	2.7	.0	100.0
		Grade Medium	5.6	94.4	.0	100.0
		Grade Low	.0	12.5	87.5	100.0

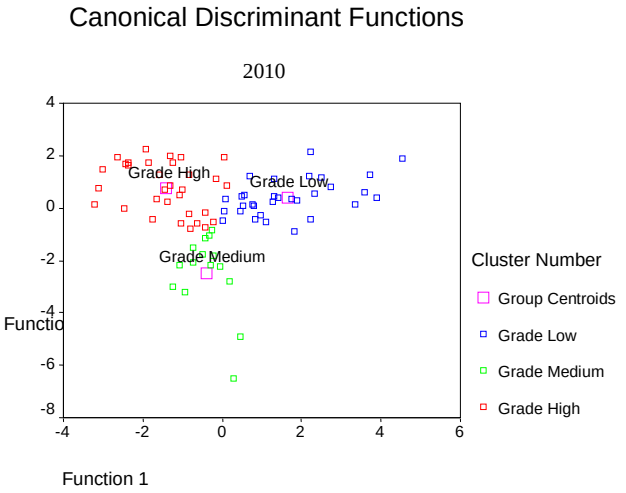
Canonical Discriminant Functions

2009



Classification Results[a]

			Predicted Group Membership			
		Cluster Number of Ca	Grade High	Grade Medium	Grade Low	Total
Original	Count	Grade High	7	0	0	7
		Grade Medium	1	52	0	53
		Grade Low	0	1	17	18
	%	Grade High	100.0	.0	.0	100.0
		Grade Medium	1.9	98.1	.0	100.0
		Grade Low	.0	5.6	94.4	100.0
Cross-validated[b]	Count	Grade High	5	2	0	7
		Grade Medium	1	51	1	53
		Grade Low	0	3	15	18
	%	Grade High	71.4	28.6	.0	100.0
		Grade Medium	1.9	96.2	1.9	100.0
		Grade Low	.0	16.7	83.3	100.0



Classification Results

			Predicted Group Membership			
		Cluster Number of Cas	Grade High	Grade Medium	Grade Low	Total
Original	Count	Grade High	31	1	0	32
		Grade Medium	0	15	0	15
		Grade Low	1	0	30	31
	%	Grade High	96.9	3.1	.0	100.0
		Grade Medium	.0	100.0	.0	100.0
		Grade Low	3.2	.0	96.8	100.0
Cross-validated	Count	Grade High	29	2	1	32
		Grade Medium	0	15	0	15
		Grade Low	4	0	27	31
	%	Grade High	90.6	6.3	3.1	100.0
		Grade Medium	.0	100.0	.0	100.0
		Grade Low	12.9	.0	87.1	100.0

With the initial labels assigned to the industries in the database, the final group was achieved by performing Multivariate Discriminant Analysis. On comparing the performances of these approaches in terms of clustering the industries, about 97 % of the companies were assigned the same grade (Tables 4 and Figures 2001 to 2010).

VI. Conclusion

In the present study, multivariate techniques such as Factor Analysis, Mac Queens' k-means cluster analysis and Multivariate Discriminant Analysis (MDA) have been used to identify the performance perspective of industries. The primary objective is to identify the first **n** number of factors that explained most of the variation in the data. Initially, factor analysis is performed to extract factor scores. Principal factors extraction is performed through SPSS package on eight financial variables on the set of ten years data. Factor extraction is used to uncover the inherent structures among the ratios considered and the central point of factor analysis is to extract factor scores for further study. The factor analysis exposed consistently three factors during each of the study period from 2001 to 2010. The amount of total variation explained by the 3 factors is over 70 percent. The factor rotations are tried with more than one method. In all that varimax rotation provided relatively better clustering of financial ratios and is included in the present study. In order to make easy interpretation, variables have been ordered and grouped by size of loading. Based on the loadings of ratios, the three factors are labeled as Debtors, Profitability and Asset /Liquidity respectively.

Then to know the performances of the industries, the k-means clustering method is used and the companies are graded H as (high), M as (medium) and L (low) based on the group means of certain industrial parameters. With the initial labels assigned to the industries in the database, the final group was achieved by performing Multivariate Discriminant Analysis. On comparing the performances of these approaches in terms of clustering the industries, about 97 % of the factors were companies the same grade. Industrial analysts can make use of the grading of industries proposed in this paper to know the performances of companies and also for the purpose of investment.

References

[1]. Altman E. I (1968), Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, The Journal of Finance, vol.23, pp. 589-609.
[2]. Altman E. I (1983), Multidimensional Graphics and Bankruptcy Prediction: A comment, Journal of Accounting Research, Volume 21, Number 1, U.S.A
[3]. Chandrasekaran R et.al,. (2013), Indian Industrial Position on the basis of Financial Ratios: A Data Mining Approach, International Journal of Statistika nad Mathematika, ISSN: 2277-2793, Volume 7, Issue 3, pp. 54-59, 2013., India.
[4]. Chandrasekaran R et.al,. (2013),Financial Ratios Performance of Major Indian Industries to evaluate their Performances using Multivariate Analyses and Perceptual Mapping, International Journal of Scientific and Engineering Research, Volume 4, Issue 2, ISS No. 2229 – 5518, **2013,** Houston, TX77043**,** USA.
[5]. Mahmoud Ezzamel, Judith and Cecilio (1987), Financial Patterns of UK Manufacturing Companies, Journal of Business Finance and Accounting, 14(4), Winter 1987, pp. 519- 536.
[6]. Richard A Johnson and Dean W. Wichern (1992), Applied Multivariate Statistical Analysis, 3/ed, Prentice-Hall of India Private Limited, New Delhi.