Indian Industrial Position on the Basis of Financial Ratios: A Data Mining Approach

R. Chandrasekaran[1†], G. Manimannan[1*] and R. Lakshmi Priya[2#]

[1]Department of Statistics, Madras Christian College, Chennai, Tamil Nadu, INDIA.

[2]Department of Statistics, Dr. Ambedkar Government Arts College, Vyasarpadi, Chennai, Tamil Nadu, INDIA.

Corresponding Addresses:

[†]drrcmcc@gmail.com, [*]manimannang@gmail.com, [#]priyagayu2006@gmail.com

Research Article

Abstract: An attempt is made to introduce a new method of grading the top ranking industries on the basis of financial ratios. It is well known that the financial ratios are being used as a measure by researchers for many purposes. About 247 companies consist of five major industries from Indian corporate database sectors were considered for each year from 2001 to 2010. Fourteen financial ratios were carefully chosen out of frequent ratios that could give different idea of the objectives and have important meaning in the literature. The unique feature of this study is the application of factor, K-mean clustering and discriminant analyses as data mining tools to develop the hidden structure present in the data for each of the study periods. Initially, factor analysis is used to uncover the patterns underlying financial ratios. The scores from extracted factors were used to find initial groups by K-mean clustering algorithm. A few outlier industries, which could not be classified to any of the larger groups, were discarded as some of the ratios possessed higher values. The clusters thus obtained formed the basis for the further analyses as they inherent the structural patterns found by the factor analysis. 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. From the present study it was observed that a little over 75% of the total variation of the data was explained by the first four factors for each industry. These four factors revealed the underlying structural patterns among the fourteen ratios that were initially considered in the analysis. Also only three clusters could be meaningfully formed for each of the periods.

Key Words: Data mining, Financial Ratios, Factor Analysis, K-means Clustering, Discriminant Analysis.

1.0 Introduction

Among various techniques used in financial statement analysis, the ratio analysis is the most powerful tool for financial analysis. The earliest study using multivariate data analysis on failure prediction was conducted by Altman (1968) using a set of financial and economic ratios as possible determinants of corporate failure. The study is used sixty six corporations manufacturing industries comprising of corporate of bankrupt and non-bankrupt firms and twenty two ratios from five categories, namely, liquidity, profitability, leverage, solvency and activity. Five rations were finally selected for their performance in the prediction of corporate bankruptcy and derived model correctly classified 95 percent of the total sample (correctly classifying 94 percent as bankrupt firms and 97 percent as non-bankrupt firms) one year period to bankruptcy. The percentage of the accuracy declined with increasing number of years before bankruptcy. As ratios are simple to calculate and easy to understand, they have been extensively used by researchers for many purposes in recent years. Altman *et al.* (1994) reported the use of neural network in identification of distressed business by the Italian central bank. Using over 1,000 sample firms with 10 financial ratios as independent variables, they found that the classification of neural networks was very close to that achieved by discriminant analysis.

2.0 Brief Review of Literature

Financial ratio analysis involves comparing the relationships between figures in the financial statements in relative terms. Key individual financial ratios appear frequently in company annual reports, auditors' reports and internal management reports. Green (1978) stated that financial ratios have long been regarded as barometers of corporate health, being used for reporting liquidity, leverage, activity and profitability and that an investor may use financial ratios to appraise a company's performance and its future prospect of success. Chen and Shimerda (1981) found that financial ratios have played an important part in evaluating the financial conditions of an entity and that over the years empirical studies have repeatedly and consistently demonstrated the usefulness of financial ratios. Chandrasekaran R, e*t. al*. (2013) grading of companies has been studied, without making any assumptions with regard to the number of groups and perceptual map, which reflected the performance of companies based on certain financial ratios. Gardiner (1995) concluded that ratio analysis continues to represent one of the financial world's most powerful and versatile tools. The earliest studies on company failures and company bankruptcies were univariate in nature. The most well-known univariate model is probably the 1966 study by Beaver, which from then on, had started many other company failure prediction analyses using other statistical techniques such as the multiple discriminant analysis by Altman (1968),

3.0 Methodology

This section brings out the discussion of the database, the ratios selected and the Data Mining Techniques.

3.1. Database and Selection of variables

The financial data published by *Capital Market* (Indian Corporate Database) was considered as the database. The data mainly consists of five major industries in India under each industry there are several companies. The data had financial ratios of each company for the time period of ten years (from 2001 to 2010). Among the listed companies, number of companies varied over the study period (*Table 1*) owing to removal of those companies for which the required data are not available. In this study, 14 ratios (Table 2) were chosen among the many that had been used in previous studies. These 14 ratios were chosen to assess profitability, solvency, liquidity, and cash-equity ratio. The choice of ratios used was based on two main criteria, namely their popularity as evidenced by their frequent usage in the finance and accounting literature and that the ratios have been shown to perform will in previous studies.

3.2 Data Mining Techniques

Although data mining is a new term, the technology is not. Data Mining or Knowledge Discovery in Databases (KDD) is the process of discovering previously unknown and potentially useful information from the data in databases. In the present context data mining exhibits the patterns by applying few techniques namely, factor analysis, **k**-means clustering and discriminant rule. As such KDD is an iterative process, which mainly consist of the following steps;

Step 1: Data cleaning;
Step 2: Data Integration;
Step 3: Data selection and transformation;
Step 4: Data Mining and
Step 5: Knowledge representation

Of these above iterative process Steps 4 and 5 are most important. If clever techniques are applied in Step 5, it provides potentially useful information that explains the hidden structure. This structure discovers knowledge that is represented visually to the user, which is the final phase of data mining.

Table 1: Number of companies in the analysis before and after Data Mining

Year	Number of Companies		Year	Number of Companies	
	Before	After		Before	After
2001	247	119	2006	247	119
2002	247	119	2007	247	119
2003	247	119	2008	247	119
2004	247	119	2009	247	119
2005	247	119	2010	247	119

3.3 Factor Analysis

Factor analysis provides the tools for analyzing the structure of the interrelationships (correlations) among the large number of variables by defining sets of variables that are highly interrelated, known as factors. In the present study, factor analysis is initiated to uncover the patterns underlying financial ratio variables (*Table 2*). In factor extraction method the number of factors is decided based on the proportion of sample variance explained. Orthogonal rotations such as Varimax and Quartimax rotations are used to measure the similarity of a variable with a factor by its factor loading.

Table 2: Variables for the study

Ratios	Description
DEB_EQU	Debt - Equity Ratio
LONG_TE	Long Term Debt-Equity Ratio
CURREN	Current Ratio
FIX_ASS	Fixed Assts
INVENTO	Inventory
DEBTORS	Debtors
INTERES	Interest
PBDITM	Profit Before Depreciation Interest Tax Margin
PBITM	Profit Before Interest Tax Margin
PBDTM	Profit Before Depreciation Tax Margin
CPM	Current Profit Margin
APATM	Adjusted Profit After Tax Margin
ROCE	Return on Capital Employed
RONW	Return on Net Worth

3.4. k-Means Clustering Methods

Nonhierarchical clustering techniques are designed to group *items,* rather than *variables,* into a collection of k clusters. The number of clusters, k, may either be specified in advance or determined as part of the clustering procedure. MacQueen (1967) suggests the term k-means for describing an algorithm of his that assigns item to the cluster having the nearest centroid (mean). Generally this technique uses Euclidean distances measures computed by variables. Since the group labels are unknown for the data set, **k**-means clustering is one such technique in applied statistics that discovers acceptable classes. Thus forming the nuclei of clusters or groups as seed points exhibited in factor analysis.

3.5 Discriminant Analysis

Multivariate Discriminant Analysis is a multivariate technique using several variables simultaneously to classify an observation into one of several a priori groups, in this case, failing and non-failing companies. In the present study, discriminant analysis is used to exhibit groups graphically and judge the nature of overall performance of the companies.

4.0 Algorithms

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

Step 1: Factor analysis is initiated to find the structural pattern underlying the data set.

Step 2: k–means analysis partitioned the data set into **k**-clusters using fourteen financial ratio data as input matrix.

Step 3: Discriminant analysis is then performed with the original ratios by considering the groups formed by the **k**-means algorithm.

5.0 Results and Discussion

As mentioned in Section 3.3 Varimax and Quartimax criterion for orthogonal rotation have been used for the pruned data. Even though the results obtained by both the criterions were very similar, the varimax rotation provided relatively better clustering of financial ratios. Consequently, only the results of varimax rotation are reported here. We have decided to retain 75 percent of total variation in the data, and thus accounted consistently four factors for each year with eigen values little less than or equal to unity. *Table 3* shows variance accounted for each factors

Table 3: Percentage of variance explained by factors

Factors	Variance explained				
	2001	**2002**	**2003**	**2004**	**2005**
1	37.06	37.85	35.58	35.85	40.22
2	17.00	15.33	18.78	16.84	19.21
3	14.39	14.63	13.53	14.07	11.30
4	8.43	8.50	8.30	8.28	7.88
Total	**76.88**	**76.31**	**76.19**	**75.04**	**78.61**
	2006	**2007**	**2008**	**2009**	**2010**
1	33.24	38.42	39.06	38.14	37.43
2	16.63	13.82	15.59	18.26	16.88
3	14.47	12.49	11.29	10.17	11.61
4	13.30	11.17	9.68	9.21	9.46
Total	**77.64**	**75.90**	**75.62**	**75.78**	**75.38**

From the above table we observe that the total variances explained by the extracted factors are over 75 percent, which are relatively higher. Also, for each factors the variability is more or less the same for the study period, though the number of companies in each year, after data cleaning and selection, kept varying owing to various reasons. The financial ratios loaded in the factors are presented in *Table 4*. Only those ratios with higher loadings are indicated by (*) symbol. From the *Table 4* it is clear that the clustering of financial ratios is stable during the study period. We observed slight changes in factor loadings during the periods considered. The differences in factor loadings may be due to statistical variations in the original data.

Table 4: Financial Ratios in Rotated Factors (Year -wise)

Initials	Measures	2001				2002				2003				2004				2005			
		Factors				Factors				Factors				Factors				Factors			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
FBDTM	Cash Equity Ratio	*				*				*				*				*			
CPM		*				*				*				*				*			
FBIDTM		*				*				*				*				*			
FBITM		*				*				*				*				*			
APATM		*				*				*				*				*			
FIX_ASS	Profitability						*				*						*		*		
ROCE		*	*			*	*				*						*		*		
RONW			*				*				*						*		*		
INTERES			*														*				
LONG_TE	Financial Leverage Ratio		*	*				*				*			*				*	*	
DEB_EQU								*				*			*					*	
INVENTO	Liquidity			*	*		*		*		*		*			*					*
CURREN				*					*				*			*					*
DEBTORS													*			*					*

Initials	Measures	2006				2007				2008				2009				2010			
		Factors				Factors				Factors				Factors				Factors			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
FBDTM	Cash Equity Ratio	*				*				*				*				*			
CPM		*				*				*				*	*			*			
FBIDTM		*				*				*				*				*			
FBITM		*				*				*				*				*			
APATM		*				*				*								*			
FIX_ASS	Profitability			*				*				*				*				*	
ROCE		*		*		*		*				*	*		*		*	*		*	
RONW			*					*				*			*			*			
INTERES																					
LONG_TE	Financial Leverage Ratio		*				*				*				*				*		
DEB_EQU			*				*				*				*				*		
INVENTO	Liquidity			*	*				*		*		*			*	*			*	*
CURREN					*				*				*				*				*
DEBTORS									*												

After performing factor analysis, the next stage is to assign initial group labels to each company. Step 2 of the algorithm is explored with factor score extracted by Step 1, by conventional **k**-means clustering analysis. Formations of clusters are explored by considering 2-clusters, 3-clusters, 4-cluster and so on. Out of all the possible trials, 3-cluster exhibited meaningful interpretation than two, four and higher clusters. Having decided to consider only 3 clusters, it is possible to rate a company as Grade **H**, Grade **M** or Grade **L** depending on whether the company belonged to Cluster 1, Cluster 2 or Cluster 3 respectively. Cluster 1 (Grade **H**) is a group of companies that have high values for the financial ratios, indicating that these companies are performing well. The companies with lower values for the financial ratios are grouped into Cluster 3 (Grade **L**). This suggested that Cluster 3 is a group of companies with low-profile. Cluster 2 (Grade **M**) are those companies which perform moderately well as compared to the Cluster 1 and Cluster

3. Inspite of incorporating the results for each year, only the summary statistics are reported in *Table 5.*

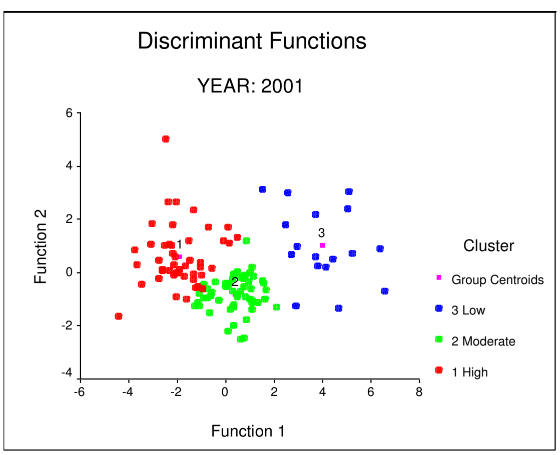
Table 5: Number of companies in the clusters

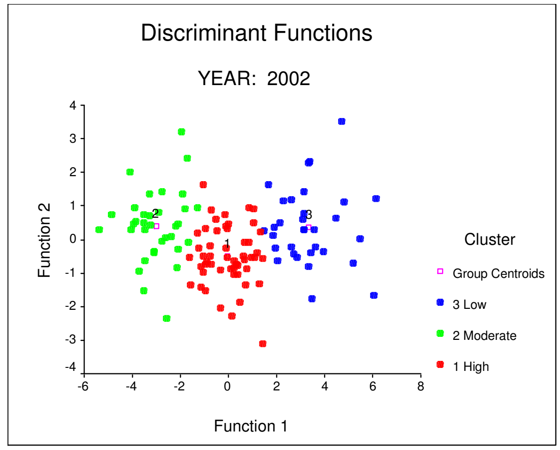
Years	**Initial Cluster**			**Discriminant Classification**		
	1	**2**	**3**	**1**	**2**	**3**
2001	17	55	47	47	55	17
2002	32	52	35	52	35	32
2003	30	86	03	04	86	29
2004	06	32	81	81	06	32
2005	55	63	01	01	55	63

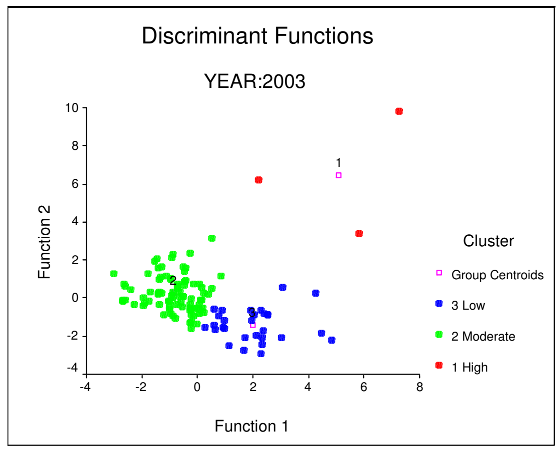
Years	**Initial Cluster**			**Discriminant Classification**		
	1	**2**	**3**	**1**	**2**	**3**
2006	10	44	65	10	65	44
2007	17	41	61	17	61	41
2008	23	58	38	58	23	38
2009	27	57	35	35	57	27
2010	44	74	01	74	44	01

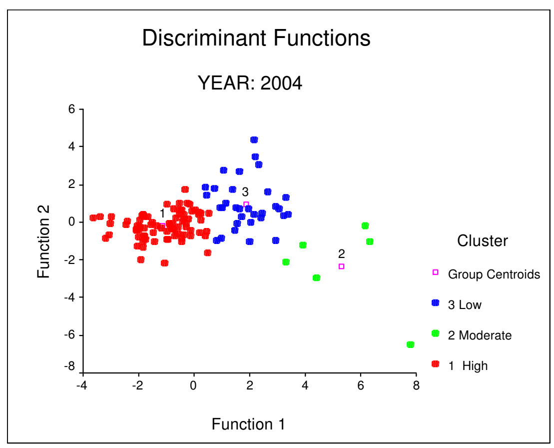
1 – Grade (**H**) 2 – Grade (**M**) 3 – Grade (**L**)

Table 5 indicates that majority of companies are in the moderate performance category except for the year 2004 and 2006. The possible reasons that kept most of the companies in lower profile in the year 2004 and 2006 may be due to the current government has concluded that most spending fails to reach its intended recipients. And also MNC's have found their way to open business in India, pushing Indian companies back. *Figure 1* through *10* shows the groupings of companies into 3 clusters for each year of the study period. It is interesting to note that the mean vectors of these clusters can be arranged in the increasing order of magnitude as show in *Table 5*.









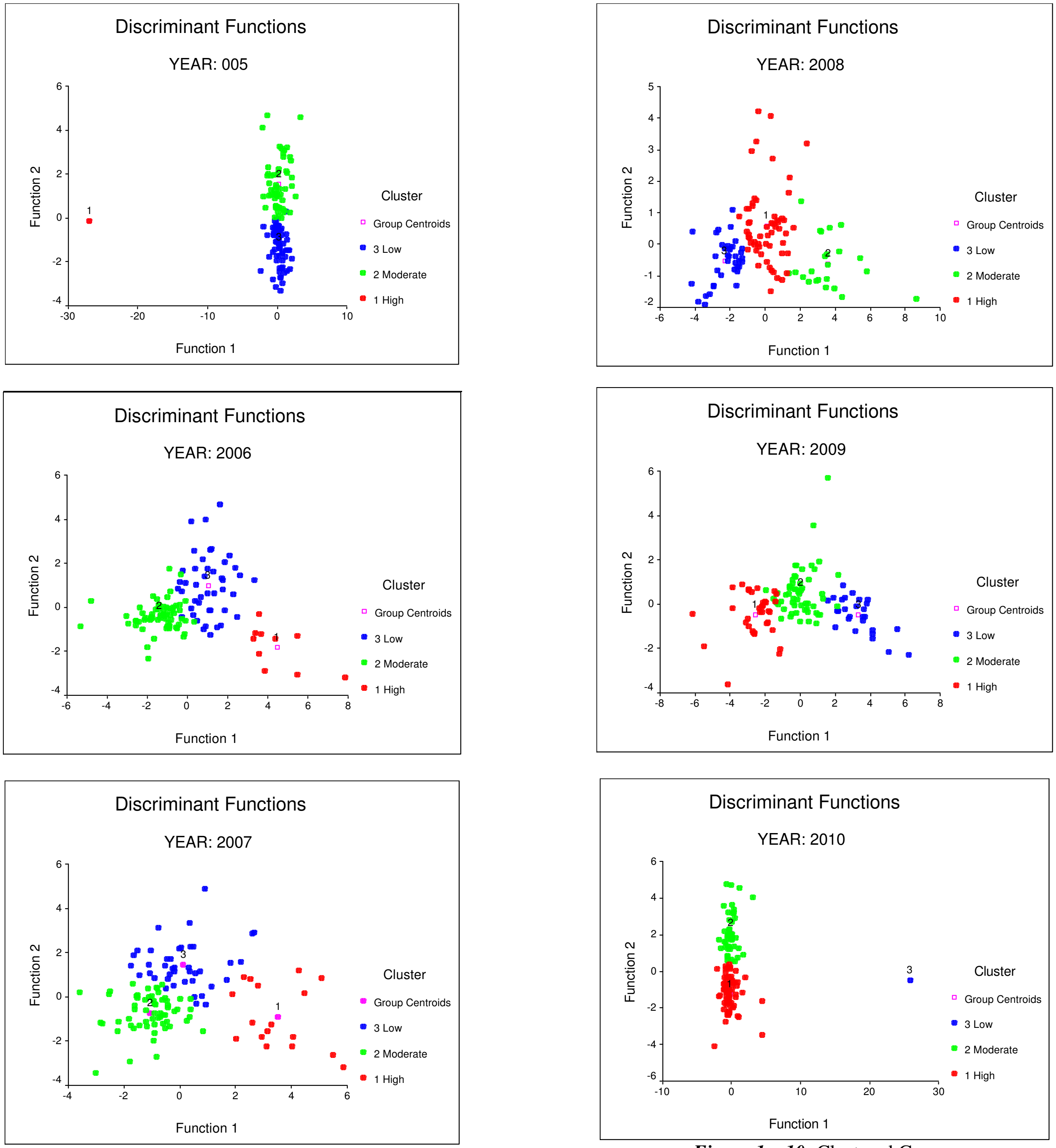


Figure 1 – 10: Clustered Groups

5.0. Conclusion

The purpose of this paper was to identify the meaningful groups of companies that are rated as best with respect to their performance in terms of financial ratios using data mining techniques. An attempt is made to analysis the financial data relating to major industries public and private sector companies over a period of ten years from 2001 to 2010. The present analysis has shown that only 3 groups could be meaningfully formed for each year. This indicates that only 3 types of companies existed over a period of ten years. Further, the companies find themselves classified into *High* (Grade **H**), *Medium* (Grade **M**) and *Low* (Grade **L**) categories depending on the financial ratios. A generalization of the results is under investigation to obtain an incorporated class of 3 groups of companies for any given year.

References

1. Anderson T W (1984), An Introduction to Multivariate Statistical Analyis, 2/e, John Wiley and Sons, Inc., New York.
2. Altman, E. I., (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, Journal of Finance, 23(4), pp. 589-609.
3. Altman, E. I., (1994). "An International Survey of Business Failure Classification Models", N.Y. University Salomon Center, Volume 6, Number 2, 1997, p 23.
4. Beaver, W. H., (1966). Financial Ratios as Predictors of Failure, Journal of Accounting Research, Vol 4. pp. 71-111.
5. Chandrasekaran, R, Manimannan G, and Lakshmi Priya, R, (2013), Financial Ratios Performance of Major Indian Industries to Evaluate their Performances using Multivariate Analyses and Perceptual Mapping, International Journal of Scientific & Engineering Research Volume 4, Issue 2, February-2013, pp, 1-6.
6. Chen, K. H. & Shimerda T. A., (1981). An Empirical Analysis of Useful Financial Ratios, Financial Management, Spring, pp. 51-60.
7. Gepp, A. and Kumar, K., (2008). The Role of Survival Analysis in Financial Distress Predictions, International Research Journal of Finance and Economics, Issue 16, pp13-34.
8. Green, D., (1978). To Predict Failure, Management Accounting July pp.39-45.
9. Li, H. & Sun, J., (2010). Business failure prediction using hybrid case-based reasoning HCBR, Computers & Operations Research January 2010. pp. 137-151.