Assessing Indian Industries on the Basis of Financial Ratios Using Certain Data Mining Tools

R. Chandrasekaran (Retd.)[1], G. Manimannan[2*], R. Lakshmi Priya[3]

{[1]Associate Professor and Head (Retd.), [2]Assistant Professor}

Department of Statistics, Madras Christian College, Chennai 600059, Tamil Nadu, INDIA.

[3]Assistant Professor, Department of Statistics, Dr. Ambedkar Govt. Arts College, Vyasarpadi, Chennai-600 039, Tamil Nadu, INDIA.

[*]Corresponding Address:

manimannang@gmail.com

Research Article

Abstract: Analyzing financial performance of companies and grading them in today's information-rich society can be a daunting task. With the evolution of the technology, Internet access to massive amounts of financial data, typically in the form of financial statements, is widespread. Managers and stakeholders of the industries are in need of the data-mining tools to quickly and accurately analyze the financial data to know about the position of their companies. An emerging technique that may be suited for such applications is the self-organizing map. The purpose of the present study is to evaluate the performances of 247 companies consist of five major industries from Indian corporate database. The time frame of the data pertaining the present study is 2001-2010. The salient feature of this study is the application of Factor, K-means clustering and Self Organizing Map (SOM) analyses as data mining tools to develop the hidden structure present in the data for each of the study periods. The scores from extracted factors are used to find initial groups by K-means clustering algorithm. A few outlier industries, which could not be classified to any of the larger groups, are discarded as some of the ratios possessed higher values. Finally, SOM is applied and the groups are identified as companies belonging to A-Class (High performance), B-Class (Moderate performance) and C-Class (Low performance) in that order. The results of the study indicate that self-organizing maps can be a feasible tool for the financial analysis of large amounts of financial ratio data.

Keywords: Financial Ratios, Data mining, Factor Analysis, K-means Clustering, Self Organizing Map (SOM)

1.0 Introduction

Analyzing financial performance of companies and grading them in today's information-rich society can be a daunting task. With the evolution of the technology, Internet access to massive amounts of financial data, typically in the form of financial statements, is widespread. Managers and stakeholders of the industries are in need of the data-mining tools to quickly and accurately analyze the financial data to know about the position of their companies. Financial parameters of banks and industries have been used in forecasting failure and bankruptcy over the past four decades. Beaver's study changed the way such analyses conducted in the field of evaluating and forecasting potential company failures and bankruptcies (Beaver, 1966). 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. Altman et al. reported the use of neural network in identifying 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. An emerging technique that may be suited for such applications is the self-organizing map.

2.0 Brief Review of Literature

Financial ratio analysis involves comparing the relationship between figures in the financial statements in relative terms. 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, productivity 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) have shown that financial ratios play an important role in evaluating the financial conditions of an entity. Further, based on their analytical studies over the years, they have demonstrated the usefulness of financial ratios. Chandrasekaran and Manimannan, *et al.* (2011) have graded companies that reflected the performance of companies based on certain financial ratios. 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 failures. The study used sixty six companies from manufacturing industries comprising of bankrupt and non-bankrupt firms and twenty two ratios from five categories, namely, liquidity, profitability, leverage, solvency and activity. Five ratios were finally selected for their performance in the prediction of corporate bankruptcy and the derived model correctly classified 95 percent of the total sample one year prior to bankruptcy. The percentage of accuracy declined with increasing number of years before bankruptcy. Altman (1994) reported the use of neural network in identification of distressed business by the Italian central bank. Using 1000 sampled firms with ten financial ratios as independent variables, they found that the neural network is not a clearly dominant mathematical model when compared to traditional statistical techniques. Other studies relating to company failure and bankruptcy using financial parameters are reported in Beaver (1966), Chen and Shimerda (1981), Gepp and Kumar (2008), Green (1978) and Li and Sun (2010).

The objective of the present study is to uncover the intrinsic groups or classes and identify the most influencing ratios that would reflect the performance of top ranking companies in India, using the concepts of Data Mining (DM), Factor Analysis (FA), Multivariate Discriminant Analysis (MDA) and Self Organizing Map (SOM).

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 types of industries in India and under each type of industry, there are several companies. The data consists of financial ratios of each company for the time period of ten years (from 2001 to 2010), around 120 companies. Among the listed companies, number of companies varied over the study period owing to removal of those companies for which the required data are not available. In this study, 14 ratios are carefully chosen among the many that had been used in previous studies *(Table 1)*. These 14 ratios are chosen to assess profitability, solvency, liquidity, and cash-equity ratio. The choice of ratios used is 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 well in previous studies.

3.2 Data Mining Techniques

Although data mining is relatively a new term, the technology is not. Data Mining or Knowledge Discovery in Databases (KDD) is the process of discovering previously hitherto 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 consists 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 the above iterative process, Steps 4 and 5 are very important. If appropriate 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.

3.2.1 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, mostly labeled, that are highly interrelated, known as factors (Anderson, 1984). In the present study, factor analysis is initiated to uncover the patterns underlying financial ratio variables (*Table 1*). 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 are used to measure the similarity of a variable with a factor by its factor loading (Everitt and Dunn, 2001; Hair, Black, Babin and Anderson, 2010).

Table 1: List of Financial parameters used in the Present study

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

3.2.2. K-Means Clustering Methods

Nonhierarchical clustering techniques are designed to group *items,* rather than *variables,* into a collection of K clusters (Everitt and Dunn, 2001; Hair, Black, Babin and Anderson, 2010). The number of clusters, K, may either be specified in advance or determined as a part of the clustering procedure. The term K-means method is coined for describing an algorithm that assigns items to the k-clusters having the nearest centroid (mean). Generally this technique uses Euclidean distance 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,i.e., groups of companies in the present study.

3.2.3 Self Organizing Map (SOM)

The self-organizing map technique creates a two-dimensional map from n-dimensional input data. This map resembles a landscape in which it is possible to identify borders that define different clusters (Kohonen, 1997). These clusters consist of input variables with similar characteristics. The methodology used when applying the self-organizing map is as follows (Back *et. al*., 1998):

Step 1. Choose the data material. It is often advisable to pre-process the input data so that the learning task of the network becomes easier (Kohonen, 1997).

Step 2. Choose the network topology, learning rate, and neighborhood radius.

Step 3. Construct the network. The construction process takes place by presenting the input data to the network iteratively using the same input vector many times, the so-called training length. The process ends when the average quantization error is small enough.

Step 4. Choose the best map for further analysis. Identify the clusters using the U-matrix and interpret the clusters (assign labels to them) using the feature planes. From the feature planes we can read per input variable per neuron the value of the variable associated with each neuron.

The network topology refers to the form of the lattice. There are two commonly used lattices, rectangular and hexagonal. In a rectangular lattice a node has four neighbors, while in a hexagonal lattice, it has six. This makes the hexagonal lattice preferable for visualization purposes (Kohonen, 1997).

The average quantization error indicates the average distance between the best matching units and the input data vectors. Generally speaking, a lower quantization error indicates a better-trained map. To visualize the final self-organizing map we use the unified distance matrix method (U-matrix). The U-matrix method can be used to discover invisible relationships in a high-dimensional data space. It also makes it possible to classify data sets into clusters of similar values. The simplest U-matrix method is to calculate the distances between neighboring neurons, and store them in a matrix. If there are "walls" between the neurons, the neighboring weights are distant, i.e. the values differ significantly. The distance values are also displayed in color when the U-matrix is visualized. Hence, dark colors represent great distances while brighter colors indicate similarities amongst the neurons (Ultsch, 1993).

3.2.4 Algorithm

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: Self Organizing Map (SOM) is then performed with the original ratios by considering the groups formed by the **k**-means algorithm.

4.0 Results and Discussion

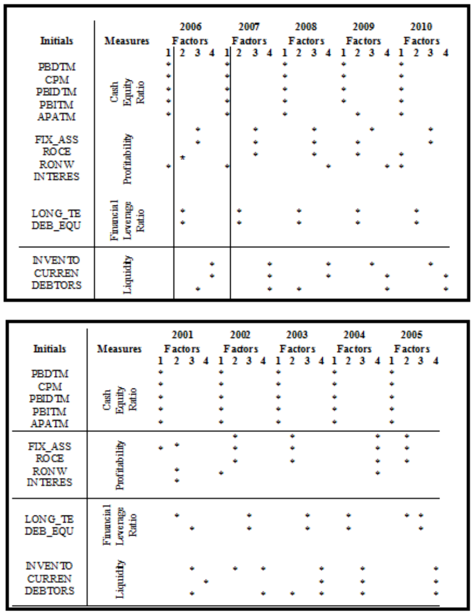
As mentioned in *Section 3.2.1,* Varimax and Quartimax criterion for orthogonal rotation have been used for the pruned data. Even though the results obtained by both the criterions were very similar, the varimax rotation provided relatively better clustering of financial ratios. Consequently, only the 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 2* shows variance accounted for each factors.

Table 2: 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 greater than 75 percent, which are relatively high. Also, for each factor 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 3*. Only those ratios with higher loadings are indicated by (*) symbol. From the *Table 3* it is clear that the clustering of financial ratios is stable during the study period. We observe 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 3: Financial Ratios in Rotated Factors



** Indicates financial ratios highly loaded in respective factors*

After performing factor analysis, the next step is to assign initial group labels to each company. Step 2 of the algorithm is applied with factor scores 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 possible trials, 3-clusters exhibited meaningful interpretation than two, four and higher clusters. Having decided to consider only 3 clusters, it is possible to classify 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 when compared with Cluster 1 and Cluster 3. In spite of incorporating the results for each year, only the summary statistics are reported in *Table4*.

Table 4: Number of companies in the clusters

Years	k-means Cluster			Self Organizing Map		
	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	k-means Cluster			Self Organizing Map		
	1	2	3	1	2	3
2006	17	55	47	47	55	17
2007	32	52	35	52	35	32
2008	30	86	03	04	86	29
2009	06	32	81	81	06	32
2010	55	63	01	01	55	63

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

Table 4 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 government policies.

And also MNC's have found their way open for business in India, pushing Indian companies back. *Figures 1* to *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 4*.

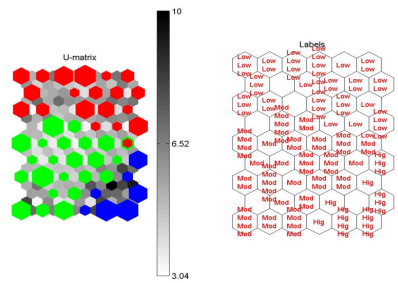


Figure 1: SOM- for the Year 2001

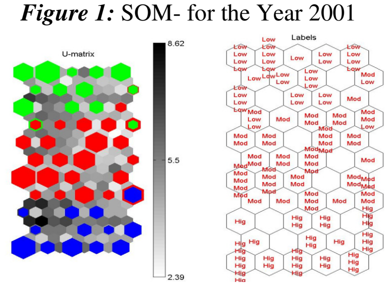


Figure 2: SOM- for the Year 2002

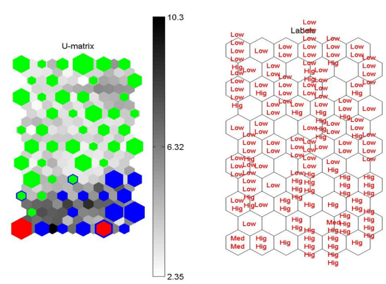


Figure 3: SOM- for the Year 2003

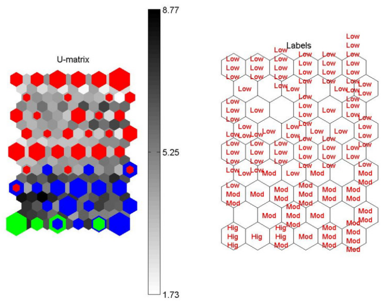


Figure 4: SOM- for the Year 2004

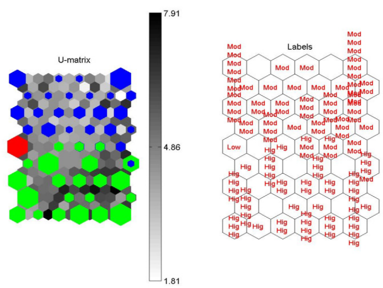


Figure 5: SOM- for the Year 2005

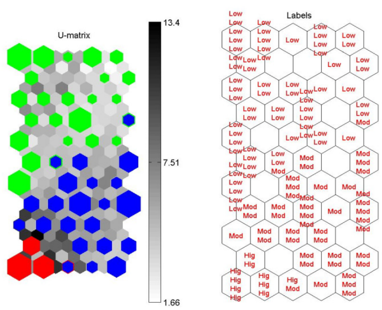


Figure 6: SOM- for the Year 2006

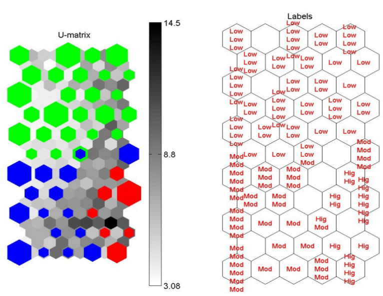


Figure 7: SOM- for the Year 2007

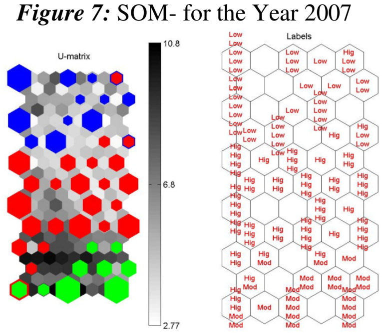


Figure 8: SOM- for the Year 2008

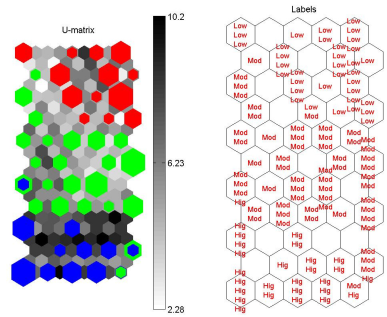


Figure 9: SOM- for the Year 2009

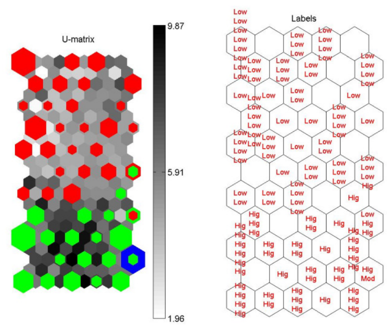


Figure 10: SOM- for the Year 2010

Inspite of incorporating the results for each year, only the summary statistics are reported in *Table 4*. The first column in *Table 4* provides the groupings done by cluster analysis. The second column gives the groupings after the application of Self Organizing Map (SOM). *Table 4* indicates that majority of companies are in the moderate performance category except for the year 2004 and 2006. Most of the companies were in lower profile in the year 2004 and 2006 may be due to the fact that the government has failed to ensure spending reach its intended recipients. And also many MNC's have found their way open for business in India, pushing Indian companies back. *Figures 1* to *10* shows the groupings of companies into 3 clusters for each year of the study period. We rated the members in the first cluster as Grade **H**, and the second as Grade **M** and the third as Grade **L**. Companies belonging to Grade **H** category are the ones that performs better than those of Grade **M** and Grade **L**. Similarly the companies belonging to Grade **M (Moderate)** category are superior to those of Grade **L**, indicating the members in the category Grade **L** are at a low profile in terms of the ratios considered in the present analysis

5.0 Conclusion

The purpose of this paper is to identify the meaningful groups of companies that are rated as best with respect to their performance in terms of financial ratios using Factor analysis, Data Mining and Self Organizing Map techniques. An attempt is made to analyze the financial data relating to major industries of public and private sector companies over a period of ten years from 2001 to 2010. The study shows 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 found themselves classified into *High* (Grade **H**), *Moderate* (Grade **M**) and *Low* (Grade **L**) categories depending on the financial ratios. Financial Analyst can make use of these techniques for rating, and the companies can project the performance on the basis of financial ratios. Self Organizing Map may be used to cross validate the performance and evaluation using financial ratios. A generalization of the results is under investigation to obtain a set 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 (2011). Assessing Indian industries on the basis of financial ratios using certain data mining tools, *Paper presented at the International Conference on Statistics and Information Analytics* (ICSIA 2012), August 23-25 2012, Department of Statistics, Loyola College, Chennai.
6. Chen, K. H. and Shimerda T. A. (1981). An empirical analysis of useful financial ratios, *Financial Management, Spring, pp. 51-60.*
7. Everitt, B. and Dunn, G. (2001). Applied Multivariate Data Analysis, 2nd Ed., Hodder Arnold, New York.
8. 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.*
9. Green, D. (1978). To predict failure, *Management Accounting, July pp.39-45.*
10. Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2010). Multivariate Data Analysis, 7th ed.,Prentice Hall, New York.
11. Li, H. and Sun, J. (2010). Business failure prediction using hybrid case-based reasoning HCBR, *Computers & Operations Research January 2010. pp. 137-151.*

Author Profile

R. Chandrasekaran received his M. Sc. M. Phil. Ph. D in Statistics from University of Madras, Chennai. He received PGDCA (Post Graduate Diploma in Computer Application).

He served as Associate Professor and Head of the Department, Department of Statistics, Madras Christian College, Chennai, India. He has good experience by working for many UGC Minor and Major Projects, Guidance for Ph. D Scholars and consultation work in application of Statistics, Pattern Recognition, Neural Networks, and Inscription. He has published more than fifty research papers in various national and International journals. He is a good expert in programming languages like, FoxPro, HTML, COBOL, C, C++, JAVA, VB, RDBMS, SPSS, SYSSTAT, STATISTICA, MINITAB, MATLAB, SAS and R Language.

G. Manimannan received his M. Sc. M. Phil. Ph. D in Statistics from University of Madras, Chennai, India. He received PGDCA (Post Graduate Diploma in Computer Application) from Pondicherry University, Pondicherry, India. He has good experience by working for many Project Guidance and consultation work in application of Statistics. He has published more than twenty research papers in various National and International journals. He is good in programming languages like, FoxPro, HTML, COBOL, C, C++, VB, DBMS, SPSS, SYSSTAT, STATISTICA, MINITAB, MATLAB and working knowledge in SAS and R.

R. Lakshmi Priya received his M. Sc. M. Phil. in Statistics from University of Madras, Chennai, India. She is Working as Assistant Professor in Statistics, Department of Statistics, Dr. Ambedkar Govt. Arts College, Vyasarpadi, Chennai. She has good knowledge in programming languages like, FORTRAN, PASCAL, COBOL, C, C++, VB and SPSS.