Performances Measures of Top-Ranking Companies Using Self Organizing Maps

G. Manimannan[1*], G. Saravanan[2], G. Senthil Kumaran[3]

[1]*Assistant Professor, Department of Mathematics, TMG College of Arts and Science Chennai, INDIA*

[2]*Assistant Professor, Department of Computer Science, St. Thomas College of Arts and Science Chennai, INDIA.*

[3] *Senior Software Engineer, GGS Information Service. India, Pvt. Ltd, Chennai*

*Corresponding Author: G. Manimannan, Assistant Professor, Department of Mathematics, TMG College of Arts and Science Chennai, INDIA

Abstract: *The research paper is to study whether the classical classification techniques, specifically the use of Self Adaptive neural network models based on Kohonen's theory, make easier to evaluate the performance of the chosen industrial sectors which are ranked according to their net sales from 2010 to 2015 with unknown group information. Out of numerous ratios that could be constructed using various financial parameters, eleven financial ratios were chosen that had different notions of the objective and significant meaning in the literature. Initially, factor analysis is applied to hidden structural patterns underlying financial ratios. The factor scores were extracted, and then used for k-means clustering techniques to prune the original database. The reduce data is then subjected to the final analysis using the proposed algorithm which results in the task of suitable grades to the companies. The method chosen also offers a way of visualization position of the company's performances through topological two-dimensional self organizing maps.*

Keywords: *Financial ratios, Factor Analysis, Data reduction, k-means clustering and Self Organizing Map.*

1. INTRODUCTION

Financial diagnosis and classification of manufacturing units has been a focal point of issue in financial analysis during the past three decades due to its severe effect on the operation of a firm, its environment and even on the whole economy of a country. Bearing in mind that the recent vulnerability of banking system and the globalization of the economy, it is apparent that financial performances of a firm is even more essential process within the financial assessment of a company by managers to place their company performances in perspective. Due to advancement in quantitative areas, such as statistics, operations research and artificial intelligence, numerous statistical classification models have been constructed for financial aspects and related problems of manufacturing companies. Several researchers, inspired by the pioneering work of Altman (1968) on the application of multivariate discriminant analysis (MDA), explored way to develop more reliable financial distress prediction model. In this context researchers typically link a set of *independent* variables to a *dependent* variable that can take two or more discrete values. Usually, only two groups are considered in financial problem where the group informations are known aprior.

It is also well known that ratio analysis is the most powerful tool for financial statement analysis that is being extensively used by researchers in building financial distress models. These include models that have been developed to predict impend corporate bankruptcy (Beaver, 1966; Altman, 1968); financial distress (Deakin, 1972; Booth, 1983, Doumpos and Zopoundis, 1999) and forewarning indicators of corporate health (Prasanth, Mishra and Satpathy, 1996). In all these models, the operational objective is to assign the firm or company to one of the group after data analysis (for example, classification of firms into *failure* versus *nonfailure*). Recently, the performances of firms are assessed using the concepts of artificial intelligence as an alternative approach. These include comparisons of corporate distress diagnosis using neural network (Altman, Marco and Varetto, 1994) and Self Organizing Map (Back, Sere and Vanharanta, 1996; Deboeck, 1998; Raquel 2001). In the present context the problem

of grading companies has been studied, without making any assumptions with regard to the number of groups or any other structural patterns in advance, which reflected the performance of Indian industries based on certain financial ratios (Chandrasekaran R and Lakshmi Priya R.). The main objective of this paper is to investigate whether;

- Data mining paradigms together with well known *unsupervised* learning neural net model *Self-Organizing Map* (SOM) can be used to exhibits the performances of top ranked companies, categorized into different industrial sectors.

- A visual representation reveals the clusters on a two-dimensional topological map.

by considering some of the categorized industrial sectors. The rest of the paper is organized as follows. Section 2 describes the methodology we have used, the database and the choice of financial ratios. Section 3 presents the proposed algorithm which is used as a target to achieve the objective on applying one of the well known neural network model SOM and Section 4 presents the empirical results. The conclusions of our study are presented in Section 5.

2. METHODOLOGY

In this section, a discussion on the database and selection of companies, the choice of financial ratios, the self-organizing principle and few data mining techniques are presented.

2.1. Database and Selection of Companies

The financial data published by *Capital Market, India* is considered for this study; the list covers public and private sector companies in India, which are rated as the best based on their net sales, from the year 2011 to 2016. Out of 1200 companies, the top most 500 companies belonging to one of the following categories viz. 1) Chemical 2) Iron and Steel 3) Textile 4) Automobiles 5) Electrical 6) Diversified 7) Cement 8) Leather and 9) Power sectors, were considered for the present analyses for each study period. Among the companies in each category, number of companies varied over the study period owing to elimination of those companies for which the required data were not available (*Table 1*).

Table1. *Number of companies in the analysis before and after Data Pruning*

Year	Number of Companies	
	Before Pruning	**After Pruning**
2011	319	315
2012	347	341
2013	351	349
2014	273	269
2015	281	268
2016	287	267

2.2. Financial Ratios

Financial ratios are simple and easy to understand. Many researchers used them to analyze some of the aspects of the firms financial situation and performances. Recently, financial ratios are used to find natural groups in large databases using SOM (Back, Kaisa and Vanharanta, 1996). As the number of ratios that can be calculated from a typical set of financial statements is much too large, only eleven financial ratios are considered that had different notions and meaningful analysis. The lists of different ratios computed and used in the present context are given in Appendix.

2.3. Self-Organizing Map

SOM is an unsupervised learning procedure based on artificial neural network. It is very effective and frequently used network popularly known as Kohonen's neural network. These networks have only two layers, a standard input layer and an output layer known as the Competitive (Kohonen) layer. Each input neuron is connected to each and every neuron on the competitive layer which are organised as a two dimensional grid. Each company is associated with exactly one neuron whereas each neuron may have one or more companies attributed to it. This grid map enables to discover statistical regularities in its input space and develops different modes of behaviour to represent different classes of input databases
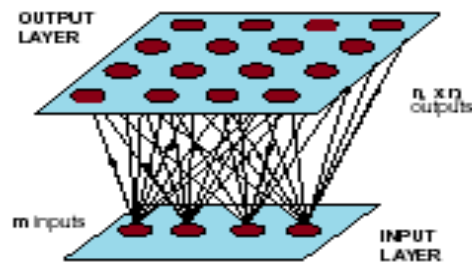
Fig1. *SOM network*

2.4. Data Mining Techniques

Data mining regarded as *the non-trivial extraction of implicit, previously unknown and potentially useful knowledge from data has* been popularly treated as synonym to Knowledge Discovery in Databases (KDD). Although data mining is a new term, the technology is not. In general, a knowledge discovery process mainly consists of iterative sequences as shown in *Figure 2*.
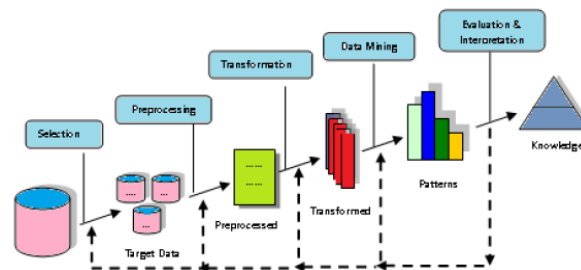

Fig2. *KDD Process*

Mining also enables the company owners to determine the impacts of sales, customer's satisfaction and corporate profits to place their company's performances in perspective. The data mining and knowledge presentation processes are the most important steps in mining process, which reveal new and hitherto unknown structural patterns present in the data. In the present context, data mining exhibits the structural patterns by applying techniques namely, factor analysis, **k**-means clustering and Multivariate Discriminant Analyses (MDA). Such structures identified from the data are presented to the user, which is the final phase of data mining.

2.4.1. Factor Analysis

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). In the present context, the application of factor analysis is centered on the parameter in the factor model that estimated values of the common factor, called factor scores using principal component method. These scores are subjected to further analyses to prune the data.

2.4.2. K-Means Clustering Algorithm

The companies considered under different industrial sectors do not have predefined labels that describe their group attribution with respect to financial status. In the present study, a nonhierarchical clustering algorithm suggested by MacQueen (1967), also known as unsupervised classification is chosen so as to group the companies considered. However, in recent days many data mining analysts make use of clustering techniques specifically in classification problems. This technique uses Euclidean distance measure computed on variables to partition or group the data set into mutually exclusive groups such that the members of each group are as close as possible to one another, and different groups are as far as possible.

2.4.3. Multivariate Discriminant Analysis

Multivariate Discriminant Analysis (MDA) is a statistical technique to classify an object into one of several a prior groupings depending upon the observations. It is used extensively by many researchers to build classification and/or make predictions when the dependent variables appear in qualitative form. In the present study, iterative discriminant analysis is used to identify the final groups and judge the nature of overall performances of the companies. This process re-allocates the companies that were assigned a group label by k-means clustering method at the initial stage. Re-allocation is subjected, allowing at the most one percent of misclassification, by considering the classification of group obtained in iteration t as the input into the next iteration t+1. It is to be noted that the concept of performing repetitive MDA is new in assessing the performances of the top ranking companies in terms of net sales.

3. ALGORITHMS

3.1. Reduced Method

In order to remove the outliers, a method to prune the data for each of the study period (during 2011 to 2016) is described below:

Step 1: Factor analysis is initiated to find the structural pattern underlying the data set and factor scores are extracted.

Step 2: k–means analysis is used to partition the data set into **k**-clusters using factor scores as input matrix.

Step 3: Repeat **Steps 1** and **2** until meaningful groups are obtained, by removing outliers in each cycle, where an outlier is a group with only a few companies in it.

3.2. Projected Algorithm

For the pruned data set the following algorithm is proposed to grade the companies and visualize them on a two-dimensional map during each of the study period based on their overall performances (Table 1).

Step 1: A map of weight vectors with 15 x 10 neurons having hexagonal topology for neighbourhood is obtained using SOM Toolbox..

Step 2: Perform k-means analysis once to partition the weight vectors obtained from Step 1 and get the group means.

Step 3: Apply k–means analysis to partition the data set, with group means of weight vectors obtained from Step 1 as initial cluster centers, and assign initial labels to the companies.

Step 4: Perform Discriminant analysis by considering the groups formed in Step 3.

Step 5: Repeat Step 4, from one iteration t to the next iteration (t+1), until at the most one percent of misclassification is obtained. This step obtains the final group means and grades are assigned to companies.

Step 6: Construct a SOM using the prototype vector with appropriate hits of the companies in the data set that are assigned group labels in Step 5.

4. RESULTS AND DISCUSSION

The results of factor analysis are initiated to extract factor scores and the data sets are combined on applying pruning algorithm in 3.1 for different values of **k**. The formations of clusters are explored by considering 2-clusters, 3-clusters, 4-clusters and so on. Isolated groups with few companies are discarded from the analysis as outliers. Out of all the possible assessments, **3**-cluster exhibited significant analysis than two, four and higher clusters. Nevertheless, it is interesting to note that factor analysis revealed consistently four factors each year that explained a little over 90 percent of total variation in the data with eigen values little less than or equal to unity.

Also, from this analysis we observed that the clustering of financial variables is stable during the study period. Although slight changes are encountered which are due to statistical variations in the original data. The clustering of financial ratios is grouped as follows; *profitability ratio* (X_2, X_3, X_4 and X_5), *capital intensiveness* (X_6, X_7, X_8, X_9 and X_{10}), *interest coverage ratio* (X_{11}) and *financial policy ratio* (X_1).

The pruned data set is then subjected to the main algorithm as in *Section 3.2* to assign appropriate grades to the companies. Initially, a SOM was trained separately, with the sequential procedure algorithm, using the SOM Toolbox version 2 for MATLAB. In the construction process several maps are initialized and trained by considering the Gaussian neighborhood function and a map with hexagonal topology. Among the prototype vector maps, the best ones in respect of average quantization error are carefully chosen to explore the data in different dimensions. From the study it is found that a weight vector map units of size 15 by 10 neurons is well spanned within the data set as in *Figure 3* for each of the study period. Similar figures are also obtained for each of the study period.

Finally, the researcher decided to consider only **3** clusters with regard to previous research affirmed in *Section 1*, the number of clusters remain unchanged for all the periods of present study. The prototype vector extracted has 150 neurons is then subjected to **k-**means algorithm with the number of classes **3**. By considering the centroids of the groups obtained from the weight vectors, the next stage in the process is to assign initial group labels to each company followed by iterative discriminant analysis in

succession allowing at most one percent of misclassification for each of the study period. The following *Table 2* summarizes the results of our study.

Table 2. *Number of companies in the clusters*

Years	Initial Cluster (1)			Converged Clusters (2)			Classification Percentage (3)	Number of Cycle (4)
	1	2	3	1	2	3		
2011	69	207	61	60	205	51	99.5	4
2012	96	235	27	93	229	22	99.1	2
2013	87	181	86	79	179	80	99.1	3
2014	130	70	129	110	65	110	99.4	4
2015	124	54	118	121	51	105	99.1	4
2016	75	173	67	65	170	69	99.2	2

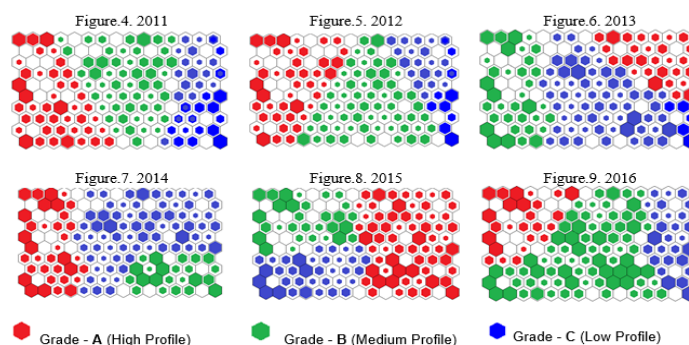1 – Grade **A**   2 – Grade **B**   3 – Grade **C**

The first column in *Table 2* provides the groupings done by **k-**mean cluster analysis considering the centroids of the respective year weight-vectors as initial centers. The second column gives the groupings after the application of discriminant analysis by allowing at the most one percent of misclassification. Column four indicates the number of cycles required for convergence. The three groups pave way to rate the members in the first cluster as Grade **A**, and the second as Grade **B** and the third as Grade **C**.

Table 2 indicates that majority of companies are in the moderate performance category except for the year 2014 and 2015. The possible reasons that kept most of the companies in lower profile in the year 2013 may be due to the political uncertainty in New Delhi and also the elections. However, the following years have seen a improvement in terms of performances for those industrial sectors considered in the study and as a result they are exhibited in 2014 and 2015 in higher profile categories. On the other hand, we observe that a considerable decrease in Grade **A** companies for the year 2016 which could be official to the ban on Rs500 and Rs1,000 notes could finally wipe off corruption in India.

In addition, SOM is used efficiently in data visualization due to its ability to represent the input data in two dimensions. Among the various visualization techniques the most widely used method for visualizing the cluster structure of the SOM is the distance matrix technique, especially the unified distance matrix (U-Matrix). For the present study, the method of displaying the number of hits in each map unit is well-thought off. *Figures* 4 through 9 show the groupings of companies into 3 clusters for each year over the SOM grid using the visualization method.

In the following *Figure*, each colour represents a grade of a company. We also observed that the mean vectors of these clusters can be arranged in the escalating order of magnitude. Companies belonging to Grade **A** category are the ones that performs better than those of Grade **B** and Grade **C**. Similarly the companies belonging to Grade **B** category are superior to those of Grade **C**, indicating the members in the category Grade **C** are at a low profile in terms of the ratios considered in the present analysis.

Self-Organizing Map of Companies (Year-wise)

Figure.4. 2011   Figure.5. 2012   Figure.6. 2013

Figure.7. 2014   Figure.8. 2015   Figure.9. 2016

Grade - A (High Profile)   Grade - B (Medium Profile)   Grade - C (Low Profile)

5. CONCLUSION

The use of this research paper is to discover the chance to identify the significant groups of companies that are rated as the best with respect to their performance in terms of net sales using SOM and few related classification techniques. An attempt is made to analyze the financial ratio data relating to public and private sector companies over a period of six years from 2011 to 2016. Initially, factor analysis is used to identify the underlying structure based on eleven financial ratios. The factor scores are used to partition the companies into different clusters by using **k-**means clustering algorithm as a methodology to reduce the data. Each year-pruned data is mined by choosing random weight vectors in a map of size

15 x 10 neurons with different training parameters and a final map with the least quantization error is obtained using SOM Toolbox. Grading of companies is attempted based on assured financial ratios applying the proposed algorithm. The capability of SOM provides several methods of representing the high dimensional vectors to be projected to a low dimension preserving the inter sample distances as possible and essential topological properties. In the present study, unified distance matrix with corresponding hits of the data to the prototype vectors are illustrated in the above *Figures 4* through *9*. This enables us to visualize the clustering of companies grouped into different categories of industries as mention in *Section 2.1* on a two dimensional map unambiguously. Hence, the members of Cluster 1 are labeled as high Grade **A** companies. Similarly, the Cluster 2 includes companies which performed moderately grade as **B** and the Cluster 3 with low profile companies with Grade **C**. 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 six years. Further, the companies find themselves classified into *High* (Grade **A**)*, Medium* (Grade **B**) and *Low* (Grade **C**) categories depending on certain financial ratios.

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] Anderson T W (1984), An Introduction to Multivariate Statistical Analysis, 2/e, John Wiley and Sons, Inc., New York.

[3] Barbro Back *et.al* (1996), Data Mining Accounting Numbers Using Self Organizing Maps, Finnish Artificial Intelligence Society, pp. 35-47.

[4] Breaver W H (1966), Financial Ratios as Predictors of Failure, Journal of Accounting Research, pp. 179-192.

[5] Chandrasekaran and Lakshmi Priya R (2014), Data Mining Application for the Performance of Indian Industries using Financial Ratios, International Journal of Computer Applications,) International Conference on Communication, Computing and Information Technology. ISSN: 0975 – 8887.

[6] David Wishart (2001), K-Mean Clustering with Outlier Detection Mixed Variables and Misssing Values, www. clustan.com

[7] Doumpos M and Zopounidis (1999), A Multicriteria Discrimination Method for the Prediction of Financail Distress: The Case of Greece, Multinational Finance Journal, Vol.3, no.2, pp.71-101.

[8] Johnson R. A and Wichern D. W. (1992), Applied Multivariate Statistical Analysis, 3/ed, Prentice-Hall of India Private Limited, New Delhi/Juha Vesanto (1999), SOM – Based Visualization Methods, http://www.cis.hut.fi/projecs/ide/publications/.

[9] Juha Vesanto (2000), Using SOM in Data Mining. http://www.cis.hut.fi/projecs/ide/publications/.

[10] Pieter Adriaans and Dolf Zantinge (1996), Data Mining, Addison-Wesly Longman Limited, England

[11] Prasanna Chandra (1997), Financial Management Theory and Practice, 4/ed, Tata McGraw-Hill Publication Company Limited, New Delhi

APPENDIX

X_1	- Debt - Equity Ratio	DEB_EQU
X_2	- Profit Before Depreciation Interest Tax Margin	PBDITM
X_3	- Profit Before Interest Tax Margin	PBITM
X_4	- Profit Before Depreciation Tax Margin	PBDTM
X_5	- Adjusted Profit After Tax Margin	APATM
X_6	-Long Term Debt-Equity Ratio	LONGTE
X_7	- Current Ratio	CURREN
X_8	- Retained Profit / Total Assets	RE
X_9	- Net Profit / Net Worth	RONW
X_{10}	- Sustainable Growth Rate	SGR
X_{11}	- Return on Capital Employed	ROCE