International Journal of Scientific and Innovative Mathematical Research(IJSIMR)
Volume 11, Issue 1, 2023, PP 1-12
ISSN No. (Online) 2347-3142
DOI: https://doi.org/10.20431/2347-3142.1101001
www.arcjournals.org

A Performance Position of Grading Companies Using Data Mining Approach

M. Aniji and G. Manimannan*

Department of Mathematics Apollo Arts and Science College Chennai, Kalvoy, Guduvanchery - 603108, Tamilnadu.

Corresponding Author:** ***G. Manimannan, *Department of Mathematics Apollo Arts and Science College Chennai, Kalvoy, Guduvanchery - 603108, Tamilnadu.*

Abstract: *Three different methods of assessment and projecting the performances of the top ranking companies on the basis of certain financial ratios based on data mining techniques are proposed in this paper. It is well known that statistical information on financial ratios is being extensively used by researchers for many purposes. To facilitate the present study, the financial information of public and private sector companies rated as the best with reference to net sales, published by Business Standard were considered for the period from 2018 to 2022. Out of numerous ratios, twenty financial ratios were sieved carefully that had different notions of the objectives and significant meaning in the literature. To exploit the hidden structure present in the data, data mining tools such as factor, k-means clustering and discriminant analyses are applied in succession. Factor analysis is initiated first to uncover the structural patterns underlying financial ratios. The scores from extracted factors were then used to find initial groups by k-means clustering algorithm to prune the data.*

*In Method **I,** clusters obtained by extracted factor are followed by iterative discriminant procedure with original ratios until cent percent classification was achieved for the year 2018. Data in the following years are trained using the previous year group means to obtain initial groups which, then iterated until cent percent classification is reached. In Method II**,** it is assumed that the rates of increase in the group means are constants from one year to the next, and the group means are increased by the corresponding rates, to mine the data for the years from 2019 onwards to inherent the patterns from the previous year. Method **III** mines the data using the estimated means for the clusters, assuming that the performance of the companies followed linear trend over the years. From the present study it is observed that the three different approaches on the classifications are equivalently good. Method **I** have slightly higher aggregate mean of 94.5% of correct classification than the other two approaches. It is also interesting to note that the clusters obtained by all the three methods could be arranged according to the magnitude of their group means on the ratios, thus permitting the groups to be identified on the basis of their performance. Finally, the groups were identified as companies belonging to Grade **A**, Grade **B** and Grade **C** in that order, which exhibit the behavior of **High** performance, **Moderate** performance and **Low** performance.*

Keywords: *Data Mining, Financial Ratios, Factor Analysis, k-mean Clustering, Discriminant Analysis, Linear Trend.*

1. Introduction

In the past two decades numerous statistical classification models have been constructed for financial related purposes. Statistical information on financial ratios are being used by the managers to place their companies performances in perspective. It is also well known that ratio analysis is the most powerful tool for financial statement analysis that are being extensively used by many researchers to meet their objective. These included models that have been developed to predict impending of corporate bankruptcy (Beaver, 1966; Altman, 1968); financial failure (Deakin, 1972; Booth, 1983); failing company (Blum, Marc, 1974); firm's performance (Bayldon, Woods and Zafiris, 1984) and forewarning indicators of corporate health (Prasant, Mishra and Satpathy, 1996). All these models typically link a set of "explanatory" variables to a "predictor" variable that can take two or more discrete values. In all these models the operational objective is to assign the firm or company to one of the group after data analysis (for example bankruptcy Vs. non bankruptcy, etc.,). The objectives of the present study are as follows:

1. To present three different methods of rating the top ranking companies on the basis of their performance using the concepts of data mining.

To identify the most appropriate proposed method of rating.

2. Database and Methodology

This section is devoted to a discussion of the database, the ratios selected for the analysis and the Data Mining Techniques.

The financial data considered for this study are published by *Business Standard* (Special Edition, BS1000 corporate Gaint), which covers the period 2018-2022 and combines public and private sector companies rated as best on their net sales in India. As it is either difficult to get the annual reports or to compare the financial, banking sectors and state sector corporations are meaningless, the publisher excluded them from the data. However, only top 500 companies are considered out of 1000 companies for the analysis for each study period. Among the listed companies, number of companies varied over the study period (*Table 1*) owing to amputation of those companies for which the required data are not available although they are published.

2.1. The Ratios

As ratios are simple and easy to understand, many researchers used them to analyse some of the aspects of the firms' financial condition and performances. However, the number of ratios that can be calculated from a typical set of financial statements is much large to in incorporate in this study. Moreover, due to availability of limited financial statement for company in the study only twenty financial ratios are sieved carefully out of numerous ratios that had different notions and meaningful interpretation. The different ratios computed are given in *Appendix*.

2.2. Data Mining Techniques

Data mining has been popularly treated as synonym to Knowledge Discovery in Databases (KDD). Although data mining is a new term, the technology is not. Researcher view data mining as the process of discovering previously unknown and potentially useful information such as patterns, associations and other significant structures from the data in databases. In general, a knowledge discovery process mainly consist of an iterative sequence of the following steps:

Step 1: Data Cleaning and Integration

Step 2: Data selection and transformation

Step 3: Data Mining

Step 4: Knowledge presentation

Mining also enables the company owners to determine the impacts of sales, customer's satisfaction and corporate profits to place their company performance in perspective. The data mining, and knowledge presentation processes are most important steps in mining process, which reveal new structural patterns present in the data. In the present context data mining exhibits the structural patterns by applying few techniques namely, factor analysis, **k**-means clustering and discriminant rule in Step 3. This structure discovers knowledge that is presented visually to the user, which is the final phase of data mining.

Table 1

Number of companies in the analysis before and after

Data Pruning

Year	Number of Companies	
	Before	**After**
2018	458	438
2019	491	466
2020	496	473

2021	399	349
2022	402	367

2.3 Factor Analysis

Different factor analysis methods are used to test the stability of financial patterns over time. Although there are several techniques of data reduction, factor analysis is by far the most frequently used method in financial researchers (Mahmoud, Judith and Cecilio, 1987). Like all data reduction methods, factor analysis reduces the variable space under consideration to a smaller number of patterns that retain most of the information contained in the original data matrix. In the present context, principle component analysis is first initiated to ascertain the structural patterns through a linear combination of the financial ratios of companies. However, in factor extraction method the first m number of factors that explained 93% of variance are considered as knowledgeable. Both orthogonal rotations such as Varimax and Quartimax rotations are used to measure the similarity of a variable with a factor by its factor loading. In factor analysis, the interest is centered on the parameter in the factor model that estimated values of the common factor, called *factor scores.* These scores are subjected to further analysis to mine the data.

2.4. k-Means Clustering Algorithm

Many data mining applications make use of clustering techniques in classifications problems. In this study, a nonhierarchical clustering algorithm suggested by MacQueen (1967) also known as ***unsupervised classification*** is well thought-of, as no presumption are made regarding the group structures present in the database. This process partition or group the data set into mutually exclusive group such that the members of each groups are as close as possible to one another and different groups are as far as possible from another. Generally this technique uses Euclidean distances measures computed by variables. The **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. The number of cluster **k** is determined to be **3** as part of the clustering procedure, which had meaningful interpretations.

2.5. Discriminant Analysis

Multivariate Discriminant Analysis (MDA) has been extensively used by many researchers in financial problems with prior group information for classification and model building. In the present study, iterative 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. Re-allocation is subjected until cent percent classification is attained, 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 DA is new in accessing the performance of the top rated companies in terms of net sales.

3. PROPOSED ALGORITHMS

A brief algorithm to prune the data during each of the study period to remove the outliers that could not be assigned to any of larger group is described below:

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

Step 2: **k**–means analysis partitioned 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.

3. 1 METHOD – I

For the pruned data set (*Table 1*) the following algorithm is proposed under this method to grade the companies during the study period 2019-2022 based on their overall performances is described below:

Step 1: Discriminant analysis is performed with original ratios by considering the groups formed by **k** –means algorithm.

Step 2: Groups means are extracted for the year **(i + 2018)** by repeating **Step 1** from iteration **t** to the next iteration **(t+1)** for some **t** until cent percent classification is achieved. **(i = 0)**

Step 3: **K** –means analysis assigns initial labels to **(i + 2019)** **th** years using group mean obtained from Step **2** as initial cluster centers.

Step 4: Repeat **Step 1** to **Step 3** for next **i**

3. 2 METHOD – II

Assuming that the rates are increase from one year to next in the financial ratios of companies are constant. A linear regression line is fitted to the final sorted group means of the pruned data sets, which are assigned initial classes by conventional **k** –mean analysis and then followed by iterative discriminant analysis. Hence, the proposed algorithm to grade the companies on the basis of overall performance from 2019 – 2022 in this method is described below:

Step 1: Increment the corresponding ratios by a constant for (**i+2018**) **th** group means respectively. **(i=0)**

Step 2: Mac Queen's **K** –means analysis assigns initial labels to **(i + 2019)** **th** years using the group mean from Step **1** as initial cluster centers.

Step 3: Discriminant analysis in then performed for (**i+1995**) **th** year by considering the groups formed by **Step 2**.

Step 4: Repeat **Step 3** until cent percent classification is achieved from iteration **t** to the next iteration (**t+1**) for some **t** and extract the group means.

Step 5: Repeat **Step 1** to **Step 3** for next **i**.

3. 3 METHOD – III

In this method, we assume that the performances of the companies over the years followed linear trend, to mine the data from 2019-2022 in order to gauge the companies according to their performances. However, in this approach Step **1** and Step **5** are excluded from method **II**, Step **2** Step **3** and Step **4** are identical for mining. The initial cluster centers are estimated from the so fitted model.

4. RESULTS AND DISCUSSION

As discussed in Section 2.3, both Varimax and Quartimax criterion of orthogonal rotation have been used for the pruned data consolidated by general algorithm. The results obtained under both the methods are very similar but the varimax rotation provided relatively better clustering of financial ratios. Factor analysis revealed consistently five factors each year that explained 90 percent of total variation in the data with eigen values little less than or equal to unity. From this analysis we observed that the clustering of financial variables is stable during the study period.

After performing factor analysis, the next stage in data mining process is to assign initial group labels to each company followed by iterative discriminant analysis in succession by the three different suggested methods. Having decided to consider only 3 clusters as mentioned in section 2.4 with regard to previous research affirmed in section 1. Inspite of incorporating the results for each method for the study periods processed through the proposed algorithms, only the summary statistics are reported in *Table 2*. The first column in *Table 2* provides the groupings done by **k**-mean cluster analysis by three different methods except the year 2018. The second column gives the groupings after the application of discriminant analysis until 100 percent classification is achieved. Column three indicates the number of cycles required for convergence using different approaches.

Table 2

Number of companies in the clusters

Years	Initial Cluster			Converged Discriminant			Number of Rotations
	1	2	3	1	2	3	
2018	150	241	47	87	329	22	14

Method – I							
2019	24	76	336	23	104	339	10
2020	17	71	385	32	123	318	20
2021	15	49	285	14	51	283	3
2022	12	45	310	12	49	306	3
Method –II							
2019	21	68	377	24	134	308	17
2020	7	66	400	22	110	341	13
2021	11	48	290	11	40	298	6
2022	10	47	310	13	47	307	6
Method –III							
2019	10	53	403	24	144	298	22
2020	17	71	385	32	123	318	20
2021	21	60	268	19	62	268	5
2022	12	45	310	12	47	308	3

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

Table 2 indicates that majority of companies are in the low performance category. *Figure 1* through *5* shows that the groupings of companies into 3 clusters for each year of the study span by method- **I**. Similarly, *Figure 6* through *11* and *Figure 12* through *17* depicts the grouping of companies into 3 clusters by method- **II** and **III** respectively. From *Table 2* it is to be noted that all the three different methods classified the companies equivalently good. However, on comparing the performances of the different approach in terms of clustering the companies as data mining process, Method **-I** had slightly higher correct classification of 93.5% than the other two methods with 92.3 % and 93.4% respectively.

From the present study we also observed that the mean vectors of these clusters can be arranged in the increasing order of magnitude as show in *Table 3-7* for each of the study periods by all the approaches. Thus, permitting is to rate the members in the first cluster as Grade **A**, and the second as Grade **B** and the third as Grade **C**. 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. Also, Method **III** performances and **I** are almost the same, therefore with any of these two methods could be used to gauge the companies performances in near future.

Table3. *Centroids of Final Groups (2019)*

Ratios	**METHOD –I**			**METHOD -II**			**METHOD -III**		
	A	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
PAY_OUT	.7564	.7482	.7146	.7609	.7609	.7074	.7475	.7262	.7127
PBDIT/GS	.1774	.1704	.1474	.1895	.1541	.1530	.1942	.1684	.1457
PBDT/GS	.1619	.1530	.0946	.1776	.1282	.0982	.1796	.1381	.0922
PAT/NS	.1214	.1067	.0684	.1185	.0891	.0724	.1327	.108	.0650
PBDIT/NS	.1922	.1887	.1656	.2153	.1714	.1698	.2391	.1889	.1594
NW/NS	.6083	.5397	.5252	.5942	.4807	.4758	.6049	.5576	.5391
CF/NS	.1120	.1086	.0822	.1079	.0937	.0863	.1167	.1040	.0803
COGS/NS	.8935	.8341	.8207	.8910	.8577	.7980	.8995	.8454	.7804
PBT/NS	.1628	.1389	.0751	.1839	.1180	.0762	.2032	.1269	.0687
PBDT/NS	.1793	.1659	.1065	.2019	.1422	.1089	.2195	.1545	.1005
PBDIT/A	.2249	.1808	.1346	.2913	.1793	.1253	.2794	.1855	.1215
PBDT/A	.2141	.1566	.0855	.2738	.1482	.0772	.2593	.1513	.0745

NS/A	1.166	1.126	1.039	1.502	1.247	.9727	1.338	1.336	.9337
PAT/A	.1439	.0942	.0539	.1627	.0898	.0502	.1594	.0927	.0477
PBIT/A	.2089	.1556	.1106	.2671	.1544	.1018	.2623	.1586	.0985
RP/A	.1118	.0715	.0406	.1223	.0691	.0375	.1186	.0707	.0360
PAT/NW	.3015	.2113	.1559	.3185	.2182	.1457	.2918	.2293	.1401
CF/NW	.2765	.2193	.1903	.2912	.2300	.1814	.2464	.2448	.1762
SGR	.0592	.0507	.0373	.0763	.0486	.0355	.0757	.0517	.0336
TIER	20.104	7.082	2.418	17.874	6.328	2.408	.0757	.0517	.0336

Table4. *Centroids of Final Groups (2020)*

Ratios	METHOD –I			METHOD -II			METHOD -III		
	A	B	C	A	B	C	A	B	C
PAY_OUT	.2855	.2755	.2645	.2912	.2691	.2632	.2855	.2755	.2645
PBDIT/GS	.1563	.1519	.1442	.1862	.1578	.1438	.1563	.1519	.1442
PBDT/GS	.1306	.1258	.0982	.1577	.1517	.0889	.1306	.1258	.0982
PAT/NS	.0826	.0720	.0692	.1093	.0995	.0619	.0826	.0721	.0692
PBDIT/NS	.1714	.1656	.1540	.2017	.1684	.1581	.1713	.1657	.1540
NW/NS	.5849	.4351	.2685	.5405	.5278	.3940	.5849	.4351	.2684
CF/NS	.0869	.0867	.0687	.1909	.0980	.0772	.0869	.0867	.0687
COGS/NS	.8926	.8629	.8605	.9028	.8382	.8294	.8926	.8629	.8605
PBT/NS	.1202	.1147	.0726	.1471	.1399	.0657	.1202	.1147	.0726
PBDT/NS	.1395	.1370	.1074	.1706	.1617	.0972	.1395	.1370	.1074
PBDIT/A	.2792	.1832	.1240	.2517	.2085	.1244	.2892	.1832	.1241
PBDT/A	.2513	.1470	.0752	.2416	.1759	.0744	.2513	.1470	.0752
NS/A	2.024	1.571	.8605	1.694	1.303	1.029	2.024	1.571	.8605
PAT/A	.1287	.0905	.0481	.1560	.1023	.0465	.1287	.0905	.0481
PBIT/A	.2436	.1592	.0998	.2278	.1800	.1006	.2436	.1592	.0998
RP/A	.0871	.0688	.0366	.1147	.0773	.0347	.0871	.0688	.0366
PAT/NW	.2914	.2137	.1267	.2995	.2261	.1303	.2914	.2137	.1267
CF/NW	.2784	.2158	.1675	.2579	.2305	.1692	.2784	.2157	.1675
SGR	.2024	.1537	.0951	.2094	.1669	.0957	.2024	.1537	.0951
TIER	13.610	6.809	2.629	25.842	6.658	2.370	13.606	6.809	2.629

Table5. *Centroids of Final Groups (2021)*

Ratios	METHOD –I			METHOD -II			METHOD -III		
	A	B	C	A	B	C	A	B	C
PAY_OUT	.2921	.2696	.2688	.2777	.2722	.2567	.2917	.2699	.2435
PBDIT/GS	.2195	.1529	.1528	.2171	.1663	.1518	.2167	.1661	.1487
PBDT/GS	.2073	.1335	.0885	.2066	.1482	.0895	.2032	.1402	.0833
PAT/NS	.1379	.0838	.0504	.1331	.0915	.05175	.1388	.0899	.0460
PBDIT/NS	.2325	.1674	.1672	.2266	.1768	.1669	.2307	.1805	.1632
NW/NS	.7701	.5733	.5558	.7155	.5745	.5713	.7154	.7009	.5406
CF/NS	.1465	.0886	.0715	.1463	.0974	.0718	.1460	.0946	.0682
COGS/NS	.9034	.8542	.7804	.9016	.8423	.7843	.9087	.8478	.7837
PBT/NS	.1810	.1182	.0593	.1719	.1297	.0617	.1817	.1210	.0542
PBDT/NS	.2196	.1457	.0966	.2157	.1577	.0984	.2163	.1521	.0912
PBDIT/A	.2298	.1907	.1320	.2333	.2040	.1335	.2406	.1796	.1298
PBDT/A	.2171	.1667	.0748	.2217	.1818	.0778	.2255	.1526	.0714
NS/A	1.304	1.271	.9903	1.315	1.264	1.000	1.327	1.163	.9974
PAT/A	.1288	.0923	.0391	.1305	.1003	.0410	.1362	.0864	.0363

PBIT/A	.1963	.1590	.1038	.1708	.1956	.1055	.2075	.1487	.1016
RP/A	.0932	.0673	.0276	.0966	.0744	.0287	.1034	.0624	.0252
PAT/NW	.2203	.1939	.0974	.2133	.2117	.1004	.2260	.1824	.0938
CF/NW	.2161	.2106	.1446	.2293	.2238	.1452	.2269	.1985	.1428
SGR	.1564	.1410	.0682	.1583	.1555	.0700	.1682	.1323	.0650
TIER	15.952	6.895	1.987	17.103	7.996	2.136	14.612	5.876	1.874

Table6. *Centroids of Final Groups (2022)*

Ratios	METHOD –I			METHOD -II			METHOD -III		
	A	B	C	A	B	C	A	B	C
PAY_OUT	.2522	.2274	.2156	.2528	.2245	.2109	.2524	.2253	.2156
PBDIT/GS	.1940	.1697	.1505	.1976	.1659	.1504	.1928	.1697	.1509
PBDT/GS	.1738	.1622	.0803	.1772	.1584	.0803	.1722	.1622	.0811
PAT/NS	.1181	.1164	.0386	.1203	.1130	.0387	.1168	.1164	.0393
PBDIT/NS	.2107	.1849	.1649	.2142	.1816	.1648	.2098	.1849	.1654
NW/NS	.6691	.6685	.6037	.6858	.6388	.6027	.6773	.6691	.6029
CF/NS	.1274	.1184	.0666	.1300	.1159	.0667	.1255	.1184	.0673
COGS/NS	.9123	.8234	.8113	.9123	.8267	.8081	.9115	.8234	.8126
PBT/NS	.1551	.1495	.0457	.1576	.1462	.0459	.1545	.1495	.0466
PBDT/NS	.1886	.1766	.0876	.1918	.1732	.0877	.1874	.1766	.0885
PBDIT/A	.2178	.2074	.1175	.2167	.2134	.1178	.2128	.2074	.1189
PBDT/A	.1974	.1954	.0639	.2026	.1946	.0643	.1974	.1903	.0656
NS/A	1.418	1.206	.8785	1.464	1.177	.8814	1.419	1.191	.8829
PAT/A	.1268	.1192	.0297	.1281	.1190	.0300	.1268	.1153	.0309
PBIT/A	.1820	.1807	.0889	.1847	.1810	.0892	.1808	.1785	.0900
RP/A	.0974	.0934	.0199	.0993	.0936	.0201	.0974	.0902	.0209
PAT/NW	.2399	.2172	.0767	.2416	.2161	.0772	.2399	.2143	.0781
CF/NW	.2360	.2288	.1322	.2358	.2355	.1324	.2332	.2288	.1333
SGR	.1806	.1687	.0516	.1837	.1686	.0518	.1806	.1667	.0526
TIER	19.420	8.071	1.841	18.502	8.064	1.849	19.042	8.319	1.844

Table7. *Centroids of Final Groups (2019)*

Ratios	METHOD –I, II & III		
	A	B	C
PAY_OUT	.4815	.2634	.2056
PBDIT/GS	.2402	.1508	.1368
PBDT/GS	.2147	.1435	.0667
PAT/NS	.1276	.0999	.0230
PBDIT/NS	.2507	.1644	.1500
NW/NS	.7468	.5794	.3329
CF/NS	.1540	.0743	.0569
COGS/NS	9273	.8432	.7759
PBT/NS	.1720	.1361	.0266
PBDT/NS	.2240	.1567	.0726
PBDIT/A	.3076	.2141	.1114
PBDT/A	.2930	.1917	.0581
NS/A	1.928	1.141	.9141
PAT/A	.1886	.1088	.0228
PBIT/A	.2702	.1692	.0793
RP/A	.1023	.0870	.0134
PAT/NW	.3118	.2011	.0431
CF/NW	.2502	.2274	.1181
SGR	.1619	.1604	.0194
TIER	27.758	12.410	1.934

Clustered Groups (METHOD – I)

Figure: 1 Year 2018

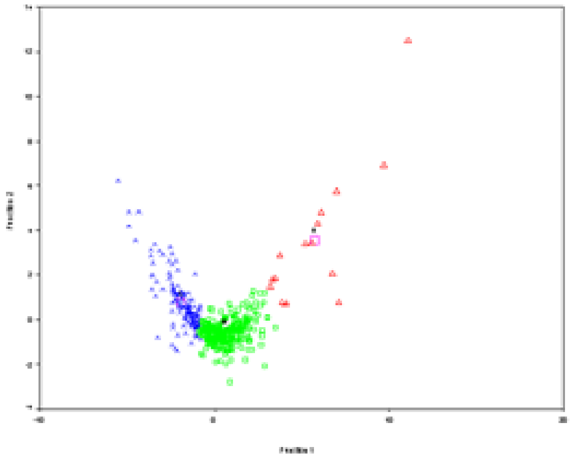


Figure: 2 Year 2019

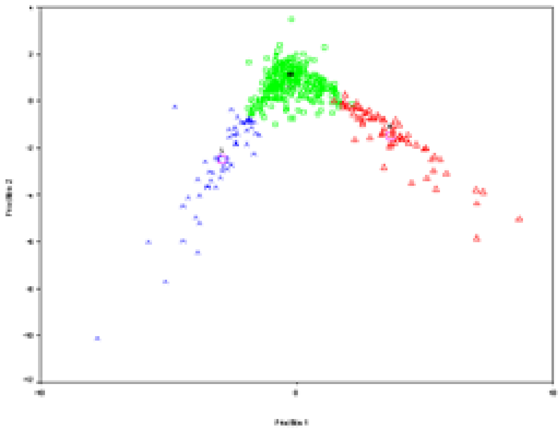


Figure: 3 Year 2020

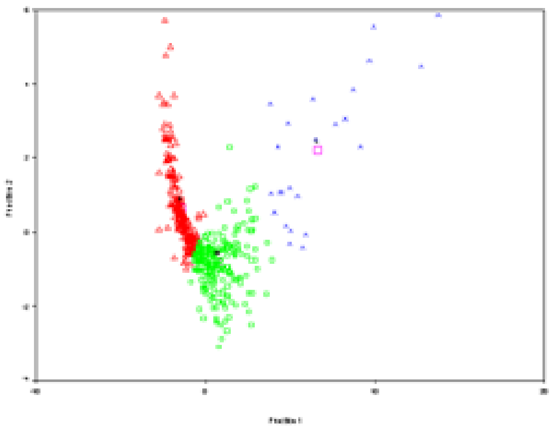


Figure: 4 Year 2021

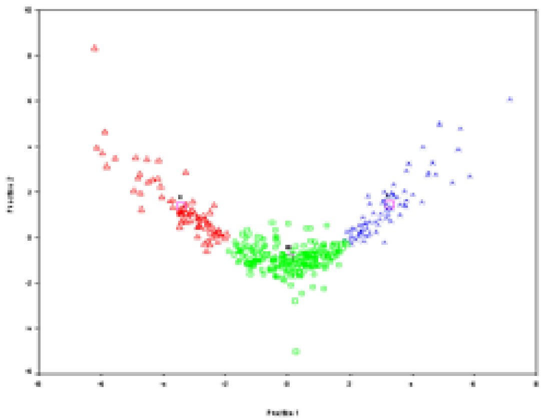
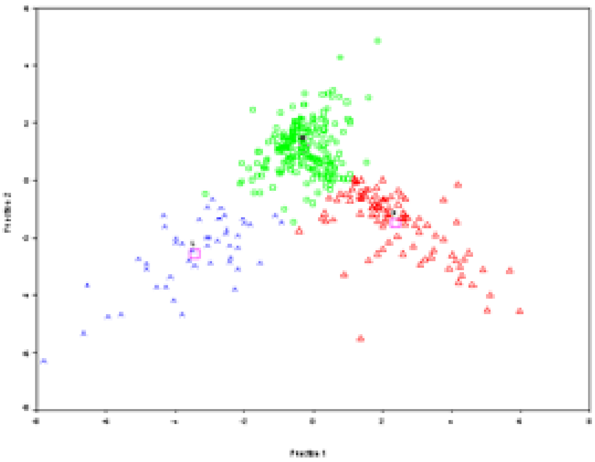


Figure: 5 Year 2022



Cluster 1 (Grade **A**) + Cluster 2 (Grade **B**) Cluster 3 (Grade **C**) Cluster centroids

Clustered Groups (METHOD – II)

Figure: 7 Year 2018

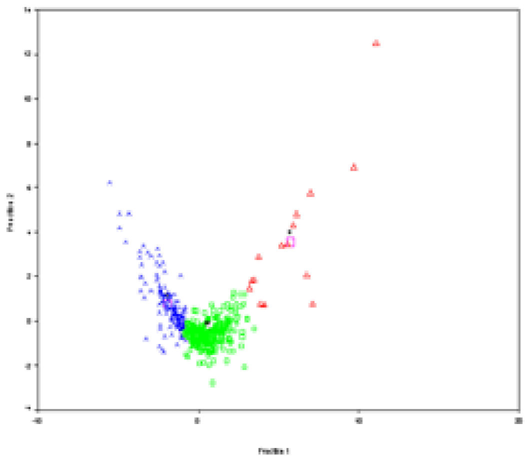


Figure: 8 Year 2019

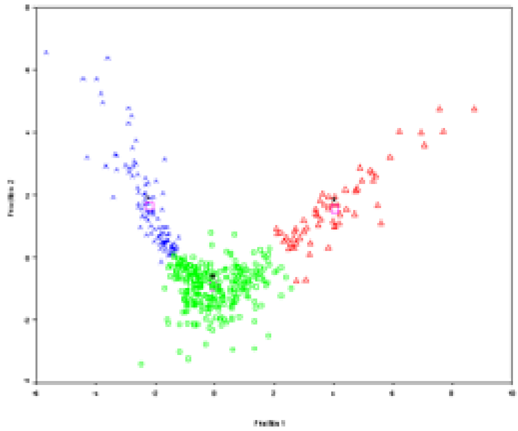


Figure: 9 Year 2020

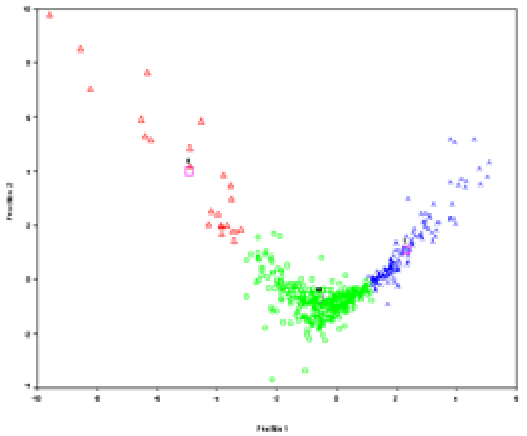


Figure: 10 Year 2021

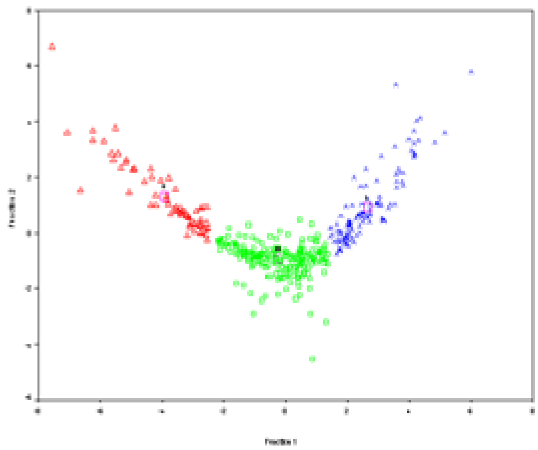
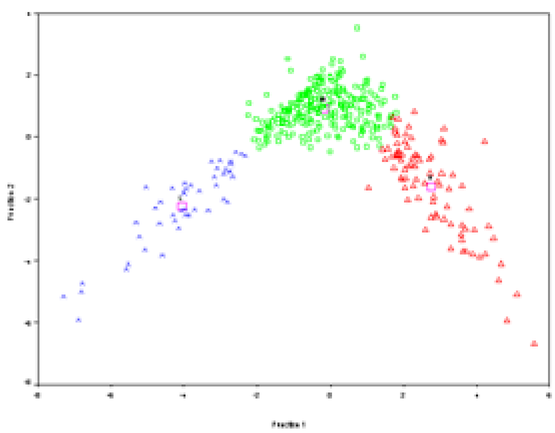


Figure: 11 Year 2022



Cluster 1 (Grade **A**) + Cluster 2 (Grade **B**) Cluster 3 (Grade **C**) Cluster centroids

Clustered Groups (METHOD – III)

Figure: 13 Year 2018

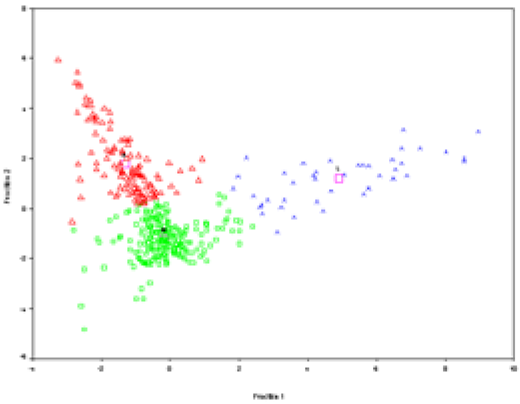


Figure: 14 Year 2019

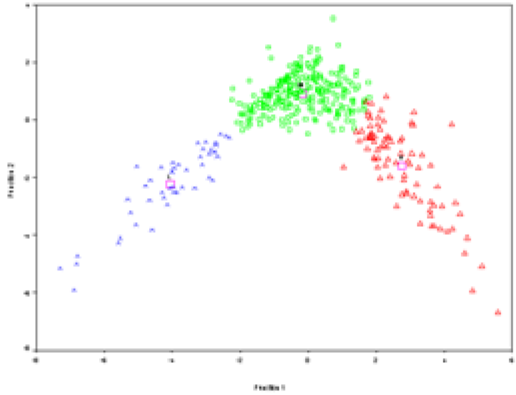


Figure: 15 Year 2020

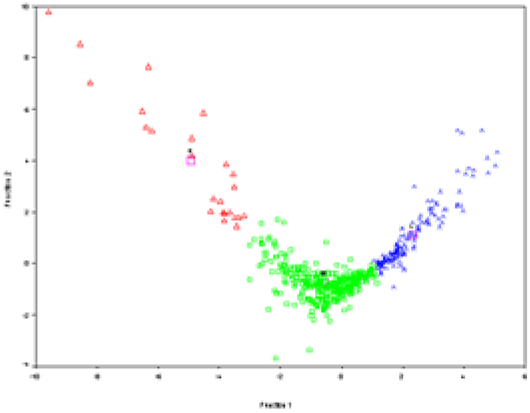


Figure: 16 Year 2021

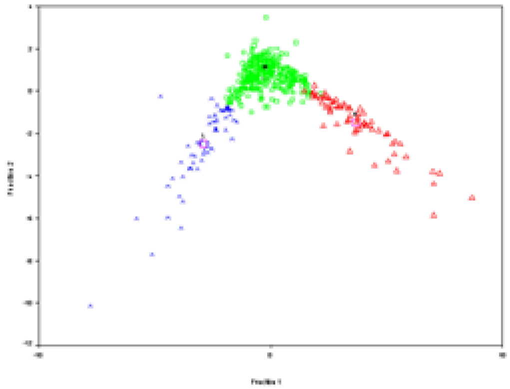
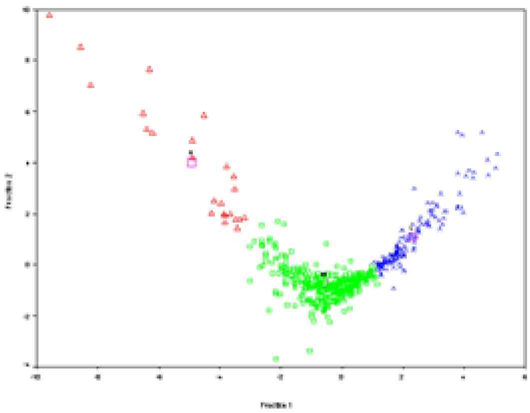


Figure: 17 Year 2022



Cluster 1 (Grade **A**) + Cluster 2 (Grade **B**) Cluster 3 (Grade **C**) Cluster centroids

5. CONCLUSION

The purpose of this paper was to propose three different methods to identify the meaningful groups of companies that are rated as best with respect to their performance in terms of net sales using data mining techniques. We attempted to analysis the financial data relating to public and private sector companies over a period of six years from2018 to 2019. Initially, factor analysis is used to identify the underlying structure in the 20 financial ratios. The factor scores are used to partition the companies into different clusters by using **k-**means clustering algorithm to prune the data.

In Method **I** successive year pruned data are mined using the previous year mean vectors as initial cluster center, keeping means vectors of 2018 as the base and so on. But the rest of the two methods mine the data by the suggested algorithm as in section 3.2 and 3.3. The unique feature in all these approach is the application of k-means as data mining tool to assign initial nuclei to cluster only once. Then they are followed by iterative discriminant analysis to re-allocate the members from iteration **t** to the next iteration **(t+1)**, until the process converges, that is, a member belonging to a cluster is assigned to itself.

The companies could be grouped only to 3 clusters for each year. The members of Cluster 1 are found to have high values for the financial ratios and hence they performed well. Thus, the members of Cluster 1 are labeled as Grade **A** companies. Similarly, the Cluster 2 included companies, which performed moderately well and the Cluster 3 with low-profile companies.

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 five years. Further, the companies find themselves classified into *High* (Grade **A**), *Medium* (Grade **B**) and *Low* (Grade **C**) categories depending on the financial ratios. A generalization of the results is under investigation to obtain a unified class of 3 groups of companies for any given year.

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 Analyis**, 2/e, John Wiley and Sons, Inc., New York

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

[4]. Blum M. (1974), Failing Company Discriminant Analysis, **Journal of Accounting Research**, pp. 179-190.

[5]. **Business Standard** (Special Edition, 1995 to 2000)

[6]. Chandrasekaran and Luther (2000), Grading Companies on the Basis of Financial Ratios: A Data Mining Approach, *Presented at the National Conference on Applied Statistics and Mathematics,* Nagpur University, Nagpur

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

[8]. David Wishart (1998), Efficiency hierarchical cluster analysis for data mining and knowledge discovery, **Computing Science and Statistics**, pp. 257-263

[9]. Deakin E. B (1972), A Discriminant Analysis of Predictor of Financial Failure, **Journal of Accounting Research**, pp. 167 – 179.

[10]. Everitt (1980), **Cluster Analysis**, Halsted Press, Division of John Wiley and Sons, New York

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

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

Appendix

Parameters used in this study

1.	Gross Profit / Net Sales	PBDT/NS
2.	Net Profit / Net Sales	PAT/NS
3.	Earning Before Interest and Tax /Total Assets	EBIT/NS
4.	Net Profit/Total Assets	PAT/A
5.	Net Profit before tax/Net Sales	PBT/NS

6.	Net Profit/Net Worth	PAT/NW
7.	Operating Profit/Net Sales	PBDIT/NS
8.	Operating Profit /Gross Sales	PBDIT/GS
9.	Gross Profit/Gross Sales	PBDT/GS
10.	Operating Profit/Total Assets	PBDIT/A
11.	Net Sales / Total Assets	NS/A
12.	Gross Profit / Total Assets	PBDT/A
13.	Cost of Sales/Net Sales	COGS/NS
14.	Cash Flow/Net Sales	CF/NS
15.	Cash Flow/Net Worth	CF/NW
16.	Net Worth/Net Sales	NW/NS
17.	Retained Earning/Total Assets	RP/A
18.	(Net profit/Net Worth) * (1- Payout)	SGR
19.	Earning Before Interest and Tax/Interest	TIER
20.	Pay out ratio	PAY_OUT

AUTHOR'S BIOGRAPHY

Dr. M. Aniji received her M. Sc. M. Phil. Ph. D in Mathematics from Anna University, Chennai, India. She received M.E. Computer Science from Anna University, Chennai, India. Currently she is working as Assistant Professor, Department of Mathematics, Apollo Arts and Science College Chennai, Tamilnadu India. She has good experience in working with many Projects Guidance and consultation work in application of Mathematical Modeling, Data mining, Neural Network, Fuzzy Logic and Other applications of Mathematics and Computer Science. She has published more than twenty research papers in various National and International journals. She has good expertise in many programming languages and software's like, FoxPro, HTML, COBOL, C, C++, JAVA, Python, PHP, Java Script, VB, RDBMS, SPSS, SYSSTAT, STATISTICA, MINITAB, Mathematica, MATLAB, SAS, R, WEKA and Orange Data Mining.

Dr. G. Manimannan received his M. Sc. M. Phil. Ph. D in Statistics from University of Madras, Chennai, India. MCA from Bharathidasan University, Trichy, India.Currently working as Assistant Professor, Department of Statistics, Apollo Arts and Science College Chennai, Tamilnadu, India. He has good experience by working with many Projects, Guidance and consultation work in application of Statistics, Data mining, Neural Network, Image Processing and Other application of Statistics and computer science. He is a co-guide of 5 research scholars, Department of Statistics, Annamalai University, Tamilnadu. He has published around hundred research papers in various National and International journals. He also published two books in Lambert Academic Publishing, Mauritius and Eastern Europe. He has good expertise in many programming languages and software's like, FoxPro, HTML, COBOL, C, C++, JAVA, Python, PHP, Java Script, VB, RDBMS, SPSS, SYSSTAT, STATISTICA, MINITAB, MATLAB, SAS, R, WEKA and Orange Data Mining.

Citation: *M. Aniji and G. Manimannan, A Performance Position of Grading Companies Using Data Mining Approach, International Journal of Scientific and Innovative Mathematical Research (IJSIMR), vol. 11, no. 1, pp. 1-12, 2023. Available : DOI: https://doi.org/ 10.20431/2347-3142.1101001*