CLUSTERING TECHNIQUES IN FINANCIAL DATA ANALYSIS APPLICATIONS ON THE U.S. FINANCIAL MARKET

ELENA CLAUDIA ŞERBAN
Ph. D. The Bucharest University of Economic Studies
e-mail: claudiaserbanos@yahoo.com
ALEXANDRU BOGEANU
PHD. Candidate The Bucharest University of Economic Studies
e-mail: alexandru.bogeanu@yahoo.com
EUGENIU TUDOR
PHD. Candidate The Bucharest University of Economic Studies
e-mail: teugeniu2003@yahoo.com

ABSTRACT

In the economic and financial analysis, the need to classify companies in terms of categories, the delimitation of which has to be clear and natural occurs frequently. The differentiation of companies by categories is performed according to the economic and financial indicators which are associated to the above. The clustering algorithms are a very powerful tool in identifying the classes of companies based on the information provided by the indicators associated to them. The last decade imposed to the economic and financial practice the use of economic value added as an indicator of synthesis of the entire activity of a company. Our study uses a sample of 106 companies in four different fields of activity; each company is identified by: Economic Value Added, Net Income, Current Sales, Equity and Stock Price. Using the ascending hierarchical classification methods and the partitioning classification methods, as well as Ward's method and k-means algorithm, we identified on the considered sample an information structure consisting of 5 rating classes.

KEY WORDS: *ANOVA analysis, clustering algorithms, rating, Economic Value Added (EVA).*

I. INTRODUCTION

Within the methods of classification, the main problem we are facing is that we do not know ″a priori ″ the number of classes we need to represent the original set of information. For example, if one's purpose is to analyze a group of companies based on the economic indexes derived from the reported accounting statements such as: company-owned equity, debt, total assets, inventory, turnover, number of employees, etc., each company from the constructed sample is displayed as identified by a 6-dimension vector; to obtain ranking among companies is a complex process which has to take into account the information contained in the analyzed variables and the number of analyzed companies. Measurement fixed errors significantly affect information and may lead to erroneous results of the performed analysis.

The economic value added (EVA). ″The economic value added index (EVA) is a privileged index for performance evaluation, which integrates the assembly of resource consumption (operational, cost of borrowed capital, and the cost of cost of the owned equity). EVA simply requires the supplementation of calculation by taking into account the cost of the owned equity.[1]"

″The recent disturbances in the financial world only increased the distrust in the „commercial" models used as a basis for measuring the performance and hierarchy of companies. A positive value of the EVA index signifies achieving wealth for shareholders in addition to the capital remuneration. A negative value shows the fact that the company fails to cover the cost of capital from the achieved operational result. In other words, the company loses money even if it reports a positive accounting result.″[2]

„Any company has to achieve, following its activity, a mean rate of return on capital markets, a rate determined under comparable risk conditions. If a company cannot generate a minimum value of return requested by shareholders, then the said shareholders will place their capital in other fields of activity or other companies[3]".

Distances between items. ″A presentation of the usual methods for calculating the distances between items (elements or already constituted groups) follows. The selection of a specific distance modifies the groups which are constituted.″[4]

---

[1] Niculescu, M.," *Global Strategy Diagnosis, Financial Diagnosis"*, Economics Publishing House, Bucharest, 2003, p.193
[2] Bogeanu, A., *″ Relevance of specific tools for an entity's performance valuation. Study of case: PETROM, a joint stock company″* Zigotto Publishing House, Galaţi, 2010, p. 36
[3] Vâlceanu, Gh., Robu, V., Georgescu, N., *Economic and financial analysis*, Economics Publishing House, Bucharest, 2005, p. 318
[4] *Multi-varied Statistics – Paper No. 12- Classification – SPSS, pp. 7,* http://profs.info.uaic.ro/~val/statistica/StatWork_12.pdf; 15.10.2012

The cluster analysis. ″The cluster analysis plays an important role within the methods of uncontrolled recognition of forms (also known as non-supervised learning methods). The purpose of the cluster analysis is represented by data classification (observations or forms) in information structures which are significant, relevant, called classes, groups or clusters.

Therefore, an essential notion used by the cluster analysis is the *cluster*. A cluster is defined as a subset of the initial aggregate of items (notes) the property of which is that the degree of dissimilarity between any of two items which belong to the cluster is lower than the degree of dissimilarity between any other item belonging to the cluster and any other item which does not belong to that cluster. It is necessary to mention a series of technical explanations. Firstly, to evaluate the distance (dissimilarity) between items (companies listed in category I) or between clusters, the Manhattan distance will be used. The Manhattan distance, also called the rectangular distance, the "City-Block" distance or norm of type L1, is calculated as a sum of the absolute values of the differences of coordinates of the two items or the two analyzed variables.

Secondly, we will use Ward's method, as a method of hierarchic classification by aggregation. This method is considered the most efficient and the highest performance method of all the "algorithms" of hierarchic classification as it the only one which explicitly treats the problem of classes homogenization, that is of minimizing of intra-cluster variability: at every step the two clusters - for which the variability of the resulted cluster is the lowest of all the possibilities of clusters fusion – are merged. An important premise of Ward's method is represented by the break-out of the total variance into intra-cluster variance and inter-cluster variance.″[5]

The nearest neighbor method. ″The distance between two groups is the minimum distance between two elements of the groups (distance between the nearest elements of different classes). This method uses the calculation of the distance between two groups as the maximum distance between two elements of the groups (distance between the farthest elements of different classes). The advantage of this method is that it does not agglomerate groups linked by a chain. Also, the achieved grouping better corresponds to the intuitive grouping (performed by a human operator).″[6]

Ward's linkage method. ″Ward's linkage is based on the growth of "sum of squares of errors" after the groups are merged into one single group. Ward's method selects the groupings which minimize the growth of sum of squares of errors.″[7]

Dendrogram. ″As a result of the algorithm, the classification ranking (dendrogram) is achieved. By the horizontal sectioning of the dendrogram, a partition of the aggregate of the classified elements is obtained. The components of the partition are the needed classes.″[8]

II. RELATED WORKS

The scientific research of a socio-economic phenomenon involves a thorough study of all the factors influencing it, and the recording of all its manifestations in the individual units. As a rule, socio-economic phenomena are highly complex and the decomposition of such a phenomenon based on its determinants becomes the subject of a complex scientific approach involving the considering the bias of the available information. The necessary appears to perform specific transformations of original data obtained by direct observation in order to extract the relevant information contained in such data. The influence of the various fields of human activity are reflected in features, attributes that describe the magnitude of their influence and allow us identify the fundamental laws which determine the dynamic evolution of analyzed the phenomena. One of the most important techniques used to synthesise information under conditions where one cannot precisely identify a functional relationship describing the dynamics of the studied phenomenon is cluster analysis. The term cluster analysis was first used in anthropological studies (Driver & Kroeber, 1932) and then in psychology (Zubin, 1938) (Tyron, 1939) and it refers, in the modern sense, to the algorithms for classifying similar items in the categories where they belong. Classification algorithms were intensively used since the development of the first numerical computers starting with mid-twentieth century. So, many applications have been developed in biology, economics, management, and sociology (Sokal & Sneath, 1963). The general action principle of the grouping techniques is to split sets of items into classes according to the analyzed attributes in such a way that each identified subset to contain similar items in terms of analyzed the characteristics. Thus, sets if items of high cardinality are represented by a few classes, thus achieving the information synthesis, facilitating the decision-making process. Classes thus identified can be used to assign a new unclassified item to the group to which it belongs by comparing its characteristics to the representative class features. In the clustering algorithms, there is no information on the classes the analyzed items belong to, this type of approach is known as unsupervised classification.

---

[5] Armeanu, Ş. D., Vintilă, G., Moscalu, M., Filipescu, M. O., Lazăr, P., (2012), ′ *Use of techniques of quantitative data analysis for the estimation of corporate bankruptcy risk′* Theoretical and applied economics, Volume XIX (2012), No. 1(566), pp. 96

[6] *Multi-varied Statistics – Paper No. 12- Classification – SPSS, pp. 7-8*, http://profs.info.uaic.ro/~val/statistica/StatWork_12.pdf; 15.10.2012

[7] *Multi-varied Statistics – Paper No. 12- Classification – SPSS, pp. 8*, http://profs.info.uaic.ro/~val/statistica/StatWork_12.pdf; 15.10.2012

[8] *Multi-varied Statistics – Paper No. 12- Classification – SPSS, pp. 9*, http://profs.info.uaic.ro/~val/statistica/StatWork_12.pdf; 15.10.2012

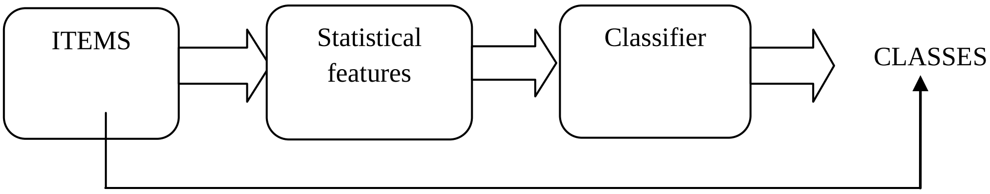


Fig. No.1 Representation of unsupervised classification process

In the unsupervised classification clustering algorithms are divided into two categories: classification algorithms by partitioning and hierarchical classification algorithms (Kaufman & Rousseeuw, 1990). The principle of action of partitioning algorithms is the segmentation of the original set of iyems in a number of classes established *a priori*. The main problem that arises in this situation is caused by the fact that there is no precise rule to determine the number of classes needed to represent the original set of items. In some applications a situation may occur when the number of classes is undervalued, that means it is considered a number of classes less than the number of classes which can be identified in reality or overrated, i.e. the number of classes is greater than the number of classes in which is actually structured the analyzed set of items and, accordingly, the information structure identified by partitioning is not representative for the analyzed statistical population.

Procedure Partitioning (D,n,k)

$P \leftarrow \emptyset;$

$for\ i = 1\ to\ k\ do\ C_i = \emptyset\ od;$

$for\ i = 1\ to\ k\ do\ C_i = RandomAssignD_i\,,\ P = P \cup D_i\ od\,;$

$while\ D \backslash P \neq \emptyset\ do\ select\ D_i \in D \backslash P\,;$

$Assign\ D_i\ to\ Cluster\ C_k\ according\ to\ Similarity\ Measure\,;$

$Recalculate\ \ Representative\ Objects\,;$

od

$repeat$

$Select\ D_i\ \in D\,;$

$Assign\ D_i\ to\ Cluster\ C_k according\ to\ Similarity\ Measure\,;$

$Recalculate\ \ Cluster\ Representative\ Object\,;$

$until\ Cluster\ Representative\ Object\ no\ Change\,;$

$for\ i = 1\ to\ k\ do\ Print\ C_k\ od;$

where : $D = initial\ \ set\ of\ objects;$

$n = number\ \ of\ objects\ contains\ in\ D, (e.g. cardinality\ of\ D);$

$k = number\ \ of\ clusters.$

From the general diagram of the partitioning algorithms one can see that, in general, as prototype item for a specific class (cluster representative item) the average cluster to a previous step can be used. Assuming that the average is statistically representative for a particular class, then the algorithm in this form is known as the „k-means algorithm” (MacKay, 2003). Other variants of partitioning algorithms use as representing items the cluster medians as, contrary to the previous situation, every time we will have as representative element an item of the set of classified items, which facilitates the logical interpretation of the formed classes. These variants that take into account class medians are grouped under the generic algorithm of „k-medoid algorithms”. Regardless of the method of selecting the representative elements for each class of clustering algorithms, the classification by partitioning requires long execution time, being included in the class of NP-hard algorithms. In terms of geometric classification partitioning algorithms resulting in obtaining classes which, in the geometric representation, are represented by convex sets. The development of calculation technique, the large amount of information stored on various media led to adjustments of partitioning algorithms leading to efficient solutions to the problem of classification. This is how the Clarans (Ng & Han, 1994) algorithm was developed. The authors propose a procedure for determining the number of clusters by running successively the proposed algorithm to determine the cluster hierarchy starting from determining the division into two clusters until the number of obtained classes is equal to the number of items to be classifiedy. The criterion for selecting the „best” number of classes is determined by calculating the silhouette coefficient, and the number of classes needed to represent the original set of items corresponds to the largest silhouette coefficient. The silhouette coefficient is determined as the ratio of the difference between the minimum measure of similarity between the item which is analyzed at some step of the algorithm and the class where it can be assigned, and the measure of similarity between the selected item and the class where it is normally assigned, and the maximum value determined between the measure of similarity associated to the current class and the minimum value of association in the rest of the classes. If this ratio is high, it means that the analyzed item is assigned to the class to which it belongs.

The hierarchical classification algorithms require to determine the number of clusters needed for the „natural" representation of the original set of items based on a special chart called dendogram (dendogram of classification). Depending on how is determined the information structure contained in the original set of items, hierarchical classification algorithms are divided into two categories: hierarchical ascending classification algorithms and descending hierarchical classification algorithms (Jain & Dubes, 1988). In the ascending hierarchical classification algorithms one starts with the distribution of each item of the original set to each individual class, then one concatenates classes based on a particular method of classification. Of the classification most used are the nearest neighbors method, the fartherst neighbors method, centroid method, the method of average distances between classes and Ward's method. In iterative stage, classes are concatenated, the classification algorithm stops when all the items are in the same class, each node of the dendogram obtained by ascending hierarchical classification is represented by a class. The convergence of ascending hierarchical algorithms is achieved in $n-1$ steps.

Procedure Ascending Hierarchical Classification (D,n)

$for\ i = 1\ to\ n\ do\ P_i = \{D_i\}\ od$

$for\ i = 1\ to\ n\ do$

$\quad for\ j = i\ to\ n\ do\ ComputeSimilarityMatrix\ \ od$

$\quad od$

$\quad k = 1$

$repeat$

$FindSimilar\ Objects\ according\ to\ Cassification\ Method$;

$Concatenate\ Calses$;

$RecalculateSimilarityMatrix\ according\ to\ Calssification\ Method$;

$k = k + 1$;

$until\ (k > n - 1)$;

where $D = set\ of\ initial\ objects$;

$\quad n = number\ of\ objects\ (e.g. carinality\ of\ D)$.

In the case of descending hierarchical classification, on its initialization, the algorithm classification starts with the class consisting of all items of the set to be classified, then, by applying the appropriate classification method one removes the item that has the dissimilarity measure maximum from the rest of the items. In this situation is made an initial partition of the set into two: one set containing n-1 items and another set containing one item. The elimination of items from a specific set is done by determining the difference between the average measure of dissimilarity within the class and average measure of dissimilarity between the selected item and the items of the set of items detached from it. The iterative step is repeated until the evaluated differences become negative. One selects the class that has the maximum diameter, and this is the new set, of which the next item is deleted, and then one resumes the iterations described above. One repeats the iterations described above until one obtains the partition which contains one item.

III. THE NECESSITY TO IMPOSE A NEW RATING MODEL BASED ON EVA

Economic value added as calculation methodology appeared and imposed itself in the early 90s, first in the U.S., then in Europe, as the most up-to-date financial performance index, an index which suggests precisely this principle of **durable development**. Through recent studies, it was found that the companies which apply EVA as a performance criterion in management agreements, do not go bankrupt so easily (as those who take profit as the performance criterion, as profit is easily to manipulate), moreover, their profitability increased on average by 183%.

The authors of the model (the "Stern Stewart Consulting Office") maintain that a company becomes profitable only when it also covers the opportunity cost of the company-owned equity (**EVA is a registered trademark of Stern Stewart & Co.**). In other words, a **company destroys economic value**, even if its reported

financial results are positive, when the **return on equity rate** is less than the **cost of company-owned equity** (a cost which is also expressed in percents, as the return on equity rate).
The "Stern Stewart Consulting Office" suggests that the Economic Value Added be:

EVA = Net operational result - Cost of invested capital

The "Stern Stewart Consulting Office" also shows how to calculate the cost of equity and the restatements of financial statements which they propose to be made in order to eliminate the "the distortions of conventional accounting"; thus it becomes possible to compare in time and space the financial statements reported by national and international companies, a comparison which can be made regardless of the accounting policies assumed by companies, of the attempts to "manipulate financial results which are used by managers" and the different tax rates existing in various countries.

The "Stern Stewart Consulting Office" presents about 160 adjustments to financial statements, but in general, a company must necessarily perform only 6 to 10 financial processing treatments, such as: research and design expenditure adjustment; goodwill adjustment; deferred tax adjustment ; provisions adjustment; operational leasing adjustment; interest deductibility adjustment.

"It is almost universally accepted the fact that the application of a score-function (rating) is limited to the period and economic zone based on which the model was developed, therefore it is questionable to use for decision making some score-functions belonging to other economies or periods."[9]

Based on a recent survey for the second quarter of 2012 on a total of 636 companies in 44 industries, listed on the NYSE, we found that a total of 178 companies had negative EVA (destroyed economic value or manipulated results), EVA being between $1 million and $-5731 million. The 178 companies with negative EVA destroyed economic value amounting to $ 45,204 million, given the situation where, of the 178 analyzed companies, 52 companies reported cumulative accounting losses totaling $ 16,548 million, the remaining 126 companies reported aggregate profits of $ 41050 million, although some others actually destroyed economic value. Overall, the 178 companies reported for the first half of the second semester of 2012, cumulative profits totaling $ 24,502 million, although EVA is negative (-45,204 million $), the difference between the reported results and EVA being of $ 69,706 million.

For the 636 analyzed companies, the cumulative EVA amounts to $ 254,919 million, and the reported results are $ 592,644 million; the reported results are being unjustifiably increased by $ 337,725 million, due to the applied accounting policies, non-recognition of expenses or revenues swelling. In percents, companies reported an economic profit increased by 132.48% versus the real situation.

There are companies which reported substantial profits, although EVA was negative each year. One such example is Company Time Warner in the Media field. For fiscal year 2011, Time Warner reported a Net Income totaling $ 2,886 million, but in reality fact the company did not create economic value added, but it destroyed it, as EVA amounts to $ -4,142 million (the difference between Net Income and EVA being -7028 million USD $ compared to the situation reported by the company).

For fiscal year 2012 Time Warner reported a Net Income totaling $ 3,019 million, but actually the company destroys economic value added, as EVA amounts to $ -3,248 million (the difference between Net Income and EVA being $ -6267 million compared to the Company Report).

IV. CASE OF STUDY

This analysis includes 106 companies listed with the New York Stock Exchange (NYSE) from the following fields: media 25, metal mining 21, software 22 and specialty retail 38 (data source: www.evadimensions.com). The sample size includes 106 companies, each company characterized by: equity, net income, current sales, price per share and economic value added (equity, net income, current sales, stock price, economic value added). Thus, each company will be characterized by a 5-dimension vector, corresponding to the five considered attributes. The data is recorded in the year 2012. The characterization of the company samples considered was performed by using the indexes of central tendency, variability and presentation of the achieved results by generating 1000 samples by the bootstrap technique.

Table no.1 **Group statistics by Industry**

Industry		Equity	Net Income	Current Sales	EVA	stock price
Media	Mean	7378.24	992.72	12269.37	-82.80	42.33
	Std. Deviation	13697.57	1555.82	14877.73	1154.43	71.12
	Range	53100.00	6496.00	58211.82	6209.80	367.80
	Kurtosis	2.63	1.51	2.96	4.90	21.87

[9] Vâlceanu Gh., Robu V., Georgescu N. - *Economic and financial analysis*, "Economics" Publishing House, Bucharest, 2005, pp 382

	Skewness	1.87	1.39	1.80	-1.54	4.55
Metals & Mining	Mean	3706.85	384.23	7210.10	79.21	30.36
	Std. Deviation	4904.55	872.56	7744.07	634.30	19.11
	Range	15948.00	4011.00	24731.60	3433.55	70.41
	Kurtosis	1.70	13.09	0.19	11.95	-0.21
	Skewness	1.69	3.47	1.16	2.89	0.56
Software	Mean	6985.68	1770.18	6880.77	1132.16	51.01
	Std. Deviation	16382.93	5233.78	16627.27	3923.33	38.13
	Range	68568.00	23453.00	72748.02	18054.65	128.96
	Kurtosis	10.73	15.17	13.09	15.85	0.06
	Skewness	3.30	3.82	3.575	3.90	0.97
Specialty Retail	Mean	3084.92	458.29	13942.38	154.22	34.14
	Std. Deviation	4372.31	1052.65	18884.87	537.40	18.39
	Range	18106.00	6887.00	70946.33	3929.07	77.19
	Kurtosis	5.32	5.56	3.25	8.86	-0.16
	Skewness	2.38	0.73	1.97	-0.24	0.66

Source: Our calculation

The companies doing business in the media field are characterized by high values of the average turnover, a high average value of the price of one share and also by the fact that the corresponding EVA is negative. The variables considered for the companies in this field are characterized by leptokurtosis of distributions and asymmetric distributions with positive asymmetry, except the EVA variable which shows negative asymmetry (skewness -1.54). For the media field, for the 21 surveyed companies, the average equity is $ 7,378.24 million, the average sales is $ 12,269.37 million, the average net income is $ 992.72 million and the average stock price is $ 42.33. However the companies in this field have negative EVA, the average being $ -82.80 million, which means that the companies in this field destroy economic value.

Companies doing business in the metal mining field are characterized by high values of the average turnover, lower value of the price of one share and by the fact that EVA is positive. The variables considered for the companies in this field are characterized by leptokurtosis of distributions and asymmetric distributions, and they all have positive asymmetry. For the metal mining field, for the 25 surveyed companies, the average equity is $ 3,706.85 million, the average sales is $ 7,210.10 million, the average net income is $ 384.23 million, and the average stock price is $ 30.36. The companies in this field differ from those in the media, by having an EVA which shows a accentuated asymmetry (most of them create economic value), the average being $ 79.21 million.

Companies that operate in the software field are characterized by lower values of the average turnover, but as they have the highest EVA, they also have the highest value of the price of one share. The variables considered for the companies in this field are characterized by leptokurtosis of distributions and asymmetric distributions and they all have positive asymmetry. For the software field, for all the 38 surveyed companies, the average equity is $ 6,985.68 million, the average sales is $ 6,880.77 million the average net income is $ 1,770.18 million, and the average stock price is $ 51.01. The companies in this field differ from those in the media, metal mining and specialty retail by the fact that EVA shows accentuated asymmetry (most of them create very high economic value added), the average being $ 1,132.16 million.

Companies that operate in the specialty retail field are characterized by the highest values of the average turnover, by high EVA, they have the average price per one share higher than those in the metal mining field. The variables considered for the companies in this field are characterized by leptokurtosis of distributions and asymmetric distributions and they all have positive asymmetry, except the EVA variable which shows negative asymmetry (skewness -0.24). For the specialty retail field, for the 38 surveyed companies, the average equity is $ 3,084.92 million, the average sales is $ 13,942.38 million, the average net income is $ 458.29 million, and the average stock price is $ 34.14. The companies in this field have EVA showing negative asymmetry (mesokurtosis), the average being g $ 154.22 million.

Table 2. **Anova Table Varaiable vs Industry**

			Sum of Squares	df	Mean Square	F	Sig.
Equity vs Industry	Between Groups	(Combined)	402530445.672	3	134176815.224	1.208	0.311
	Within Groups		11327803658.667	102	111056898.614		
	Total		11730334104.340	105			
NetIncome vs Industry	Between Groups	(Combined)	29516733.722	3	9838911.241	1.455	0.231
	Within Groups		689563547.938	102	6760426.941		
	Total		719080281.660	105			
Current	Between	(Combined)	1037488898.704	3	345829632.901	1.383	0.252

Sales vs Industry	Groups						
	Within Groups		25513164637.787	102	250129065.076		
	Total		26550653536.491	105			
EVA vs Industry	Between Groups	(Combined)	20710000.799	3	6903333.600	1.883	0.137
	Within Groups		373961808.739	102	3666292.243		
	Total		394671809.538	105			
stock price vs Industry	Between Groups	(Combined)	5913.125	3	1971.042	1.171	0.325
	Within Groups		171760.916	102	1683.931		
	Total		177674.041	105			

Source: Our calculations

The constructed sample is identical with respect to the companies found in the four examined fields of activity, which is emphasized by the fact that the differences between the variables reported for each group of companies are not statistically significant, for the five analyzed variables. This is confirmed by the rejection of the null hypothesis which considers that there are significant differences between the analyzed variables and the grouping of the companies by industry. The rejection of this hypothesis is made by determining the F statistic evaluated for each variable and field of activities, the materiality threshold of 5% being exceeded every time. Thus, we conclude that although there are major differences assessed for each industry in terms of the analyzed indexes, the differences within groups and between groups are not significant, which means that the field of activity is not a discriminating factor. The equivalence of groups of companies in terms of area of activity where they operate, determines the comparability of companies from different fields.

Table no. 3 **Descriptive statistics for Equity Variable**

Descriptives						
			Statistic	Std. Error	Bootstrap 95% Confidence Interval Lower	Bootstrap 95% Confidence Interval Upper
Equity	Mean		5030.30	1026.62	3278.44	7126.79
	95% Confidence Interval for Mean	Lower Bound	2994.71			
		Upper Bound	7065.89			
	5% Trimmed Mean		3221.66		2110.75	5211.44
	Median		1320.50		1037.00	2273.25
	Std. Deviation		10569.65		6489.32	14064.18
	Minimum		-5624.00			
	Maximum		68659.00			
	Range		74283.00			
	Interquartile Range		3581.75		2238.88	5615.13
	Skewness		3.73	0.23	2.79	4.39
	Kurtosis		16.03	0.47	8.10	21.83

Source: Our calculations

The 95% confidence interval estimated from bootstrap reselections, with its lower bound around 3278 and respectively the upper bound of $ 7.126 billion, is similar to the parametric confidence interval, estimated on the basis of mathematical selection theory from normally distributed variables (normal distribution) and this match suggests that for the 106 companies in the sample, the average value of capitals is 5030 (order of magnitude). But the probability distribution associated to this characteristic is asymmetric, oriented to values higher than the average value, as results from the skewness coefficient, which shows that median value is a better index to characterize the series of data. Also, the value of the kurtosis coefficient is large, displaying the leptokurtosis character of the distribution of the equity variable. The significant difference between the value of selection average and the value of average α-truncated (5% Trimmed Mean) is justified by the presence of high levels of capitals, the amplitude of the series being $74,283 million. The presence of outliers within these variable records justifies the use of the median value as an appropriate index for measuring the central tendency, the range of variation of the median having the lower bound 1037 and the upper bound 2273.

Table no. 4 **Descriptive statistics for Net Income Variable**

Descriptives			
	Statistic	Std. Error	Bootstrap 95% Confidence Interval

				Lower	Upper
Net Income	Mean		841.94	254.18	474.55	1359.90
	95% Confidence Interval for Mean	Lower Bound	337.95			
		Upper Bound	1345.94			
	5% Trimmed Mean		495.06		317.43	747.58
	Median		180.50		112.00	312.32
	Std. Deviation		2616.94		1045.20	4019.29
	Minimum		-2781.00			
	Maximum		23344.00			
	Range		26125.00			
	Interquartile Range		751.75		395.00	989.85
	Skewness		6.63	0.23	4.57	6.82
	Kurtosis		53.61	0.47	23.33	39.01

Source: Our calculations

After analyzing the companies in terms of their net profit variable, it is found that the confidence interval obtained for the average by generating random samples (bootstrap sampling), is similar to the parametric confidence interval, having about the same length, which –for the original sample – makes us consider that the average profit of the companies is $ 841.94 million. However, the great length of the intervals obtained for this variable and the positive asymmetry shown by the skewness coefficient causes the average to be a bias estimator to characterize the central tendency existing in the data series, associated to the net income variable. The distribution of data associated to the companies in the sample has leptokurtosis character, but the kurtosis coefficient, is outside the confidence interval obtained by bootstrap resampling, thus justifying the presence of outliers. The confidence interval for the median value is narrower and less than the range of average variation, which makes the median net income in the analyzed sample be an index that catches the central tendency better than the average of the series. Thus, we can consider that the average value of the variable net income is $ 180.5 million, with the lower bound of $ 112 million and the upper bound of $ 312.32 million, respectively, for the marginal probability of 5%.

Table no. 5 **Descriptive statistics for Economic Value Added variable**

Descriptives						
			Statistic	Std. Error	Bootstrap	
					95% Confidence Interval	
					Lower	Upper
EVA	Mean		286.43	188.31	13.34	697.26
	95% Confidence Interval for Mean	Lower Bound	-86.95			
		Upper Bound	659.81			
	5% Trimmed Mean		103.59		-8.84	251.18
	Median		20.08		5.61	70.30
	Std. Deviation		1938.76		682.00	3020.70
	Minimum		-3913.64			
	Maximum		17469.74			
	Range		21383.38			
	Interquartile Range		256.54		147.56	565.86
	Skewness		6.99	0.23	4.70	7.38
	Kurtosis		60.66	0.47	26.24	43.10

Source: Our calculations

For the EVA variable, the similarity of the confidence intervals for the average, the confidence interval obtained by bootstrap reselections is similar to the confidence interval obtained if the parametric version of it is considered, and shows that the sample companies created economic added value amounting to $ 286.43 million; however, their great length and the presence of asymmetry (skewed distribution) as the distribution of values associated to this variable determines the inconsistency of the average as an index which characterizes the data series. The confidence interval associated to the median EVA being very narrow and much smaller in terms of values than variation ranges of the average, the accentuated positive symmetry associated to the variable results in considering the median as a value representative for the group of sample companies. The statistical distribution associated to the series of data corresponding to the EVA variable has a leptokurtosis character, the kurtosis coefficient is not contained in the associated range of variation, the high amplitude thereof (range = 21383.38) is justified by the presence of outliers.

Table no. 6 **Descriptives statistics for Current Sales variable**

Descriptives						
			Statistic	Std. Error	Bootstrap 95% Confidence Interval Lower	Bootstrap 95% Confidence Interval Upper
Current Sales	Mean		10748.44	1544.51	8067.20	13839.52
	95% Confidence Interval for Mean	Lower Bound	7685.96			
		Upper Bound	13810.91			
	5% Trimmed Mean		8356.17		5967.67	11714.69
	Median		4280.19		2863.50	6395.60
	Std. Deviation		15901.68		11999.87	19437.60
	Minimum		267.40			
	Maximum		73031.00			
	Range		72763.60			
	Interquartile Range		11371.36		7358.58	17747.01
	Skewness		2.39	0.23	1.84	3.03
	Kurtosis		5.65	0.47	2.51	10.50

Source: Our calculations

The group of companies considered in the sample had an average turnover of $ 19,748,44 million, the insignificant differences between the confidence intervals assessed for a probability of 95% distribution asymmetry less pronounced than in the previous cases makes the average turnover value be a correct indicator for the sample of considered companies. The confidence intervals obtained by bootstrap resampling for the median value, the skewness coefficient and the kurtosis coefficient contain the accurate values evaluated for the sample, so that, from the distribution of the turnover values, one finds that the companies are relatively compact. Even if the range of the data series associated to the turnover is high ($ 73031.00 million), it is found that the interquartile range is close to the average value of the series, which shows the existence of balance between companies with high turnover and those with relatively low turnover.

Table no. 7 **Descriptive statistics for stock price variable**

Descriptives						
			Statistic	Std. Error	Bootstrap 95% Confidence Interval Lower	Bootstrap 95% Confidence Interval Upper
stock price	Mean		38.83	4.00	32.44	47.05
	95% Confidence Interval for Mean	Lower Bound	30.90			
		Upper Bound	46.75			
	5% Trimmed Mean		33.84		29.54	38.69
	Median		30.14		27.68	35.28
	Std. Deviation		41.14		22.04	60.55
	Minimum		5.87			
	Maximum		373.82			
	Range		367.95			
	Interquartile Range		31.05		25.05	35.80
	Skewness		5.50	0.23	1.88	5.29
	Kurtosis		42.04	0.47	16.61	33.99

Source: Our calculations

In the case of the stock price variable, the average share price for the sample of companies analyzed is $ 38.83, the confidence intervals achieved by parametric and nonparametric evaluation corresponding to the level of result guarantee of 95% were about the same length, but the positive asymmetry associated to the set of data shows the distribution orientation to values in excess of the selection average, the distribution has leptokurtosis character, the kurtosis coefficient determined for the sample being outside the associated confidence interval [16.61, 39.99]. The asymmetry of the data series, the shorter length of the confidence interval for median versus the confidence intervals associated to average shows that the data series is affected by the presence of outliers, small interquartile range shows that 50% of the considered companies do not report differences greater than $ 32 between the associated share price associated.

The descriptive analysis of the five variables considered in our review highlights the positive asymmetry associated to the distributions of data series, the presence of outliers in the data series, the lack of illustrative character of selection media, which makes the company group heterogeneous, and the differences versus the average value truncated α-(5% trimmed mean), showing the pronounced presence of outliers. The analyzed variables have relatively small interquartile range in relation to the order of magnitude associated thereto and a relatively large number of observations are distributed around the selection average, making the distributions associated to these variables have a leptokurtosis character. The high amplitudes associated to each variable show the high degree of heterogeneity of the statistical observations.

All these considerations are in favor of the hypothesis that the data series associated to the companies in the four fields of activity are not derived from the normal distribution.

Table no. 8 **Analysis of distribution**

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig. level	Statistic	df	Sig. level
Equity	0.300	106	< 0.001	0.520	106	< 0.001
NetIncome	0.309	106	< 0.001	0.387	106	< 0.001
EVA 2012	0.327	106	< 0.001	0.351	106	< 0.001
Current Sales	0.255	106	< 0.001	0.659	106	< 0.001
Stock Price	0.212	106	< 0.001	0.561	106	< 0.001

Source: Our calculations

The testing of normality of probability distributions, confirms the assumption that the analyzed data series are not normally distributed, both the Kolmogorov-Smirnov test and the Shapiro-Wilk test are statistically significant for all the variables studied for the 0.1% threshold (significance level), which means that the sample data does not fit the normal distribution. Failure to identify the distribution - which includes the analyzed statistical observations - makes it difficult to determine the causal relationships existing between the analyzed variables with respect to such relationships.

The normality of the analyzed data series is also invalidated by the representation of quantiles of the theoretical normal distribution and the quantiles of the empirical distribution which were studied. The observed values deviate from the theoretical quantiles of the normal distribution for all the analyzed variables, both for the small values of variables, and especially when they increase. The values associated to current sales are the most distanced from the corresponding values of the normal distribution, especially when Current Sales increases. The same situation is also encountered for the Equity variable, but the tendency to move relative to the theoretical values is more attenuated.

Because the distributions the original data come from do not show any normality trace, we can say that – for the entire analyzed sample - the average selection is not representative, as it is affected by the presence of outliers in the statistical observations. Therefore, the use of the median value to characterize the sample in terms of the central tendency is appropriate.

The confirmation of the use of median as an indicator of the central tendency, existing within the observations is justified by the values of the estimators -as Hampel meant this- whose values are very close to the values of selection medians.

Table no. 9 **Hample's Estimators**

M-Estimators				
		Statistic	95% Confidence Interval	
			Lower	Upper
Equity	Hampel's M-Estimator	1415.19	1003.20	2072.22
Net Income	Hampel's M-Estimator	194.38	108.513136	344.93
EVA	Hampel's M-Estimator	33.33	-2.145801	106.01
Current Sales	Hampel's M-Estimator	4438.10	2829.00	6537.97
Stock Price	Hampel's M-Estimator	31.82	27.96	35.99

Source: Our calculations

The high heterogeneity of the data which characterizes the companies in the considered sample, the large amount of information (each objects is characterized by five features), justifies the use of the cluster analysis techniques to perform information synthesis. The surprising aspect for pattern recognition techniques) is that it allows the observation of regularities in the absence of an "a priori" information about the representation of the analyzed phenomenon in the form of a mathematical model which describes its dynamics. The logical deployment of a cluster analysis aiming at the partition of a set of heterogeneous objects set by homogeneous classes in terms of the employed attributes implies the choice of features depending on which the classification is to be performed, the choice of type of measure of proximity between objects, the determination of the rules of class forming, the check of class consistency and significance.

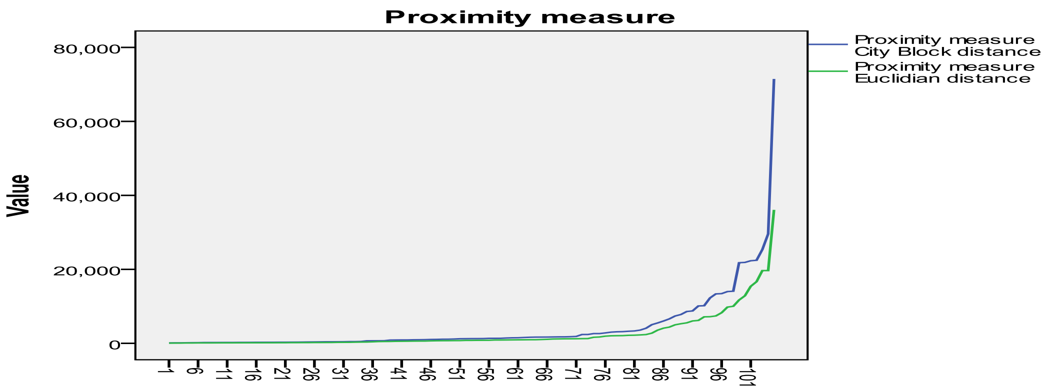
Because the characteristics associated to the objects in the original data matrix are heterogeneous –a fact which is justified by the pronounced influence of the reported outliers -, this makes the average not significant, and the use of Euclidean distance leads to inaccurate representations of these objects in the predictor variable space, therefore, in our analysis, we will use the Manhattan distance, generically called "city-block distance", which is less affected by the presence of outliers, and it does not distort the "order of the objects represented in the space of features.

The identification of outliers is performed using the nearest neighbor method which involves the evaluation of distances between the classes due to a specific iteration of the classification process, as being the distance between the "closest" objects of the two classes.

At a certain stage of classification, a new class of reunion of the classes which are the closest. Given the right rule of classes formation, at the end of the classification process, outliers will be distributed in classes, which contain only the objects in question, which makes their interpretation difficult to make.

To achieve the conclusive results, these records will be removed from the sample because they affect the decision-making process during classification because they differ significantly from other objects.

Graph 2. **Representation of proximity measures**



Source: Our calculations

Using the Manhattan distance as a measure of proximity between objects and as the method of classification, the nearest neighbor method - a group of companies consisting of Microsoft. and Oracle Corp. – differentiate in the considered sample, both being from the computer information field.

Table no.10 **Outliers identification using k-nearest neighbors**

Focal Record	Nearest Neighbors					
	1	2	3	4	5	6
MSFT	CMCSA	HD	NWS	LOW	TWX	AA
	Nearest Distances					
	1	2	3	4	5	6
	1.731	1.99	2.182	2.24	2.559	2.619
Focal Record	Nearest Neighbors					
	1	2	3	4	5	6
ORCL	NWS	DIS	FCX	LOW	TWX	AA
	Nearest Distances					
	1	2	3	4	5	6
	0.839	0.894	1.049	1.194	1.199	1.318

Source: Our calculations

Microsoft Corp – KPI is the best performing company in the Software field. For June 2012, Current Sales amount to $73,031 million, Net Income amounts to 23,344 million, Equity amounts to $ 68,659 million, EVA amounts to $ 17,740 million, and Stock Price is $ 30,59 on June 30, 2012.

Oracle Corp – KPI is the best performing company in the software field. For June 2012, Current Sales amount to $ 36,980 million, Net Income amounts to $ 9,739 million, Equity amounts to $ 42,873 million, EVA amounts to $ 6,696 million, and Stock Price amounts to $ 29.70 on June 30, 2012.

Table no. 11 **Average silhouette for different numbers of clusters**

Clusters numbers	Average silhouette	Net Income	EVA	Equity	Current Sales	Stock Price
2	0.8	0.88	1	0.3	0.11	0.72
3	0.9	0.88	1	0.32	0.13	0.01
4	0.8	1	0.96	0.56	0.16	0.57

5	0.8	1	0.82	0.65	0.48	0.4
6	0.8	1	0.98	0.89	0.58	0.42
7	0.8	0.93	1	0.87	0.56	0.38
8	0.7	0.96	1	0.88	0.62	0.36
9	0.7	1	0.96	0.85	0.6	0.35
10	0.7	1	0.99	0.8	0.63	0.3
11	0.7	1	0.95	0.75	0.57	0.27
12	0.7	1	0.96	0.73	0.56	0.26
13	0.7	1	0.96	0.75	0.56	0.25
14	0.5	1	0.98	0.7	0.58	0.53
15	0.6	1	0.91	0.66	0.53	0.49

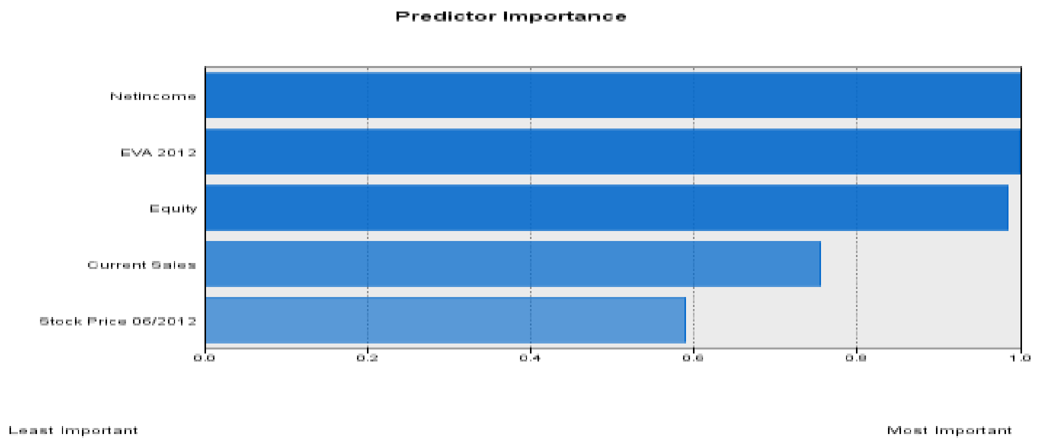
Source: Our calculations

For the performed analysis, the most important features used, i.e. those which provide maximum discrimination between the analyzed objects are Net income, EVA and Equity. Seen as a whole, the Current Sales variable has little importance in the process of classification, when the number of classes is low.

According to the Stock Price variable, the analyzed companies are different when the number of classes is small, with great contribution to the make of the classification, and they display a decreasing trend when the number of classes increases, which means that the analyzed companies –located in the same class- are identical in terms of the analyzed variable.

Overall, of the five variables associated to each company, it is found that they all have a high discriminatory power, having values above 0.5 for the average silhouette.

Graph 3. **Predictor importance in classification process**

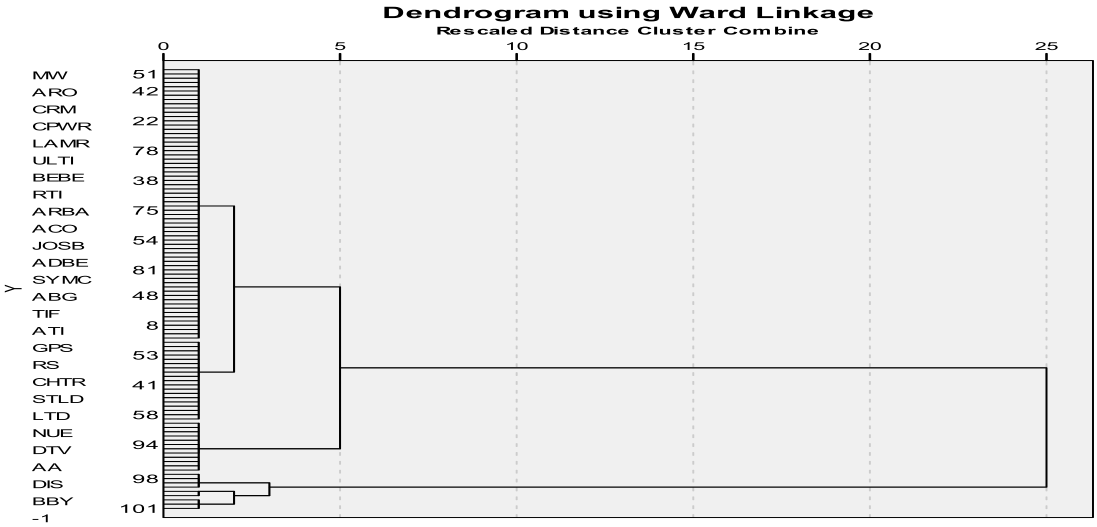


Source: Our calculations

The determination of classes is performed by using Ward's method of classification, which involves maximizing the homogeneity of classes by minimizing the variability of the measured classes as the sum of the squares of deviations in the cluster configuration.

The characteristic of minimizing the intracluster variability, is obtaining a maximum homogeneity in a given configuration of objects on clusters. After eliminating outliers values and applying the Ward method of classification, two classes are differentiated.

Graph 4. **Dendogram for Ward linkage method**



Source: Our calculations

The first class includes 95 sample companies, and the second class includes 9 companies. However, for the second class, it is found that homogeneity is reduced, which leads to its structuring into two subclasses.

Table no. 12 **Typical cluster identification**

Cluster		Equity	Net Income	Current Sales	EVA	Stock Price
1	Mean	16622.80	824.80	41349.20	-950.32	34.02
	Std. Deviation	11918.19	2695.78	9743.28	1957.81	15.92
	Maximum	29780.00	3415.00	51176.00	841.70	59.70
	Minimum	4514.00	-2781.00	29270.00	-3913.64	20.96
2	Mean	29840.75	4163.50	60569.25	380.31	47.91
	Std. Deviation	15433.20	930.53	14008.57	1596.81	11.34
	Maximum	47476.00	5170.00	71380.00	1996.07	58.19
	Minimum	15863.00	2937.00	41508.00	-1816.25	31.97

Source: Our calculations

The companies in the first class have all their indexes lower than those in class II (Current Sales, Net Income, Equity and Stock Price). The EVA in the first class is negative, average EVA being $ -950.30 million (companies in this class destroy economic value). The EVA in class II is positive, the average being $ 380.31 million (companies in this class create economic value added).

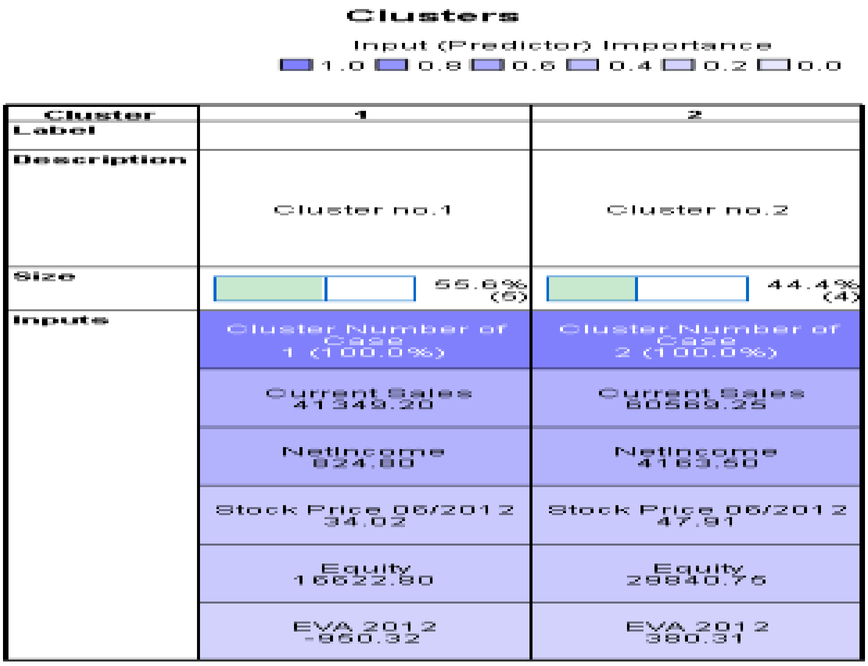
Table no. 13 **Importance predictors for clusters**

Variable	F-statistics
Equity	2.119
Net Income	5.476
Current Sales	5.934
EVA	1.198
Stock Price	2.144

Source: Our calculations

Of the five variables characterizing the companies in the achieved clusters, it is found that the most important for the discrimination between these classes are the Current Sales and Net Income variables. Analyzing the contribution to the separation of companies, it is found that in these classes there are companies characterized by high values of net income and Current Sales - between the two identified classes there are low values of EVA.

Graph 5. **Typical cluster member**



Source: Our calculations

A typical company in the second cluster is characterized by positive values of EVA ($ +380.31 million), Current Sales are 50% higher than with the companies in the first cluster, Net Income is 5 times higher than with the companies in the first cluster and Stock Price is 40.8% higher.

Applying the classification method to the 95 companies in the first class leads to the identification of five subsets of objects which organizes –in terms of information– the original set of information objects.

In terms of significance, the first 3 clusters are differentiated, which group 82 objects of the 92 objects present in the original set. For the 95 companies, we obtain the following cluster configuration.

Table no.14 **Cluster configuration using Ward method**

Ward Method						
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid clusters	1	44	46.3	46.3	46.3
	2	19	20.0	20.0	66.3
	3	20	21.1	21.1	87.4
	4	7	7.4	7.4	94.7
	5	5	5.3	5.3	100.0
	Total	95	100.0	100.0	

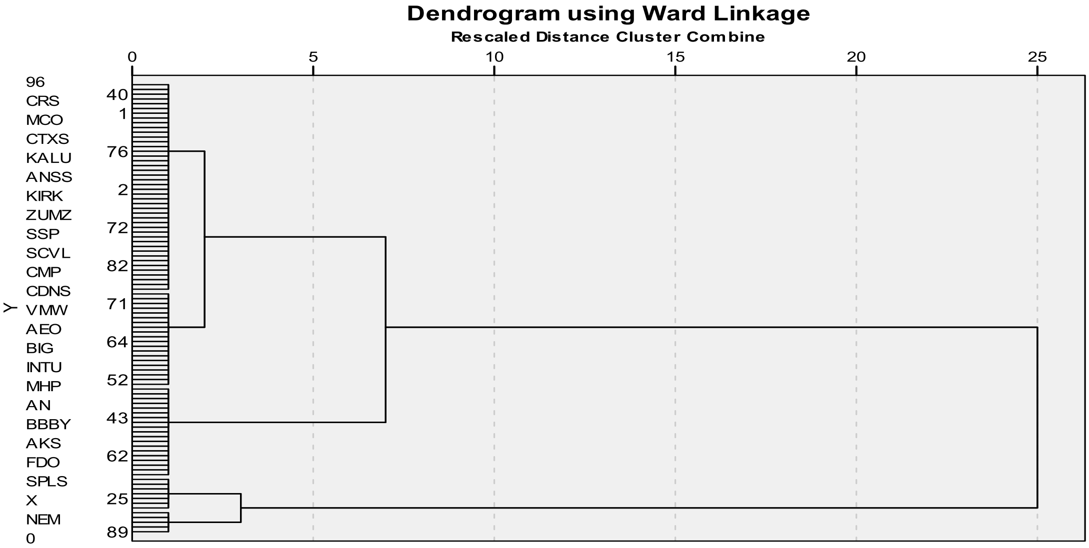
Source: Our calculations

The maximization of intra-cluster variability shows that, for representing the original aggregate of the data set, it is necessary to consider a structure consisting of 5 classes.

The elements of the identified classes are characterized by the fact that they have a maximum degree of homogeneity in terms of variability evaluated in terms of distance.

In terms of significance, the first 3 clusters are differentiated which group 82 objects of the 92 objects present in the original set, as resulting from the dendogram associated to the classification.

Graph 6 **Dendogram for classification of data after outliers analysis**



Source: Our calculations

Of the identified classes, clusters 1 and 2 are more similar and clusters 1 and 3 are less similar. Also, the elements of the first cluster are very different from all other clusters, with the largest distances to the centers of the other clusters. Cluster 3 is very close to cluster 5, which makes the objects of such clusters be similar in terms of the values of the analyzed variables. Cluster 4 is the most "isolated" of the 5 identified clusters, as it has no cluster corresponding to any other cluster; the elements contained in this cluster will be very different from the other companies.

Table no. 15 **Distances between Final Cluster Means**

Cluster	1	2	3	4	5
1	0	7198.419	22189.209	16216.302	21988.945
2	7198.419	0	15480.694	12031.992	14831.589
3	22189.209	15480.694	0	10383.792	8623.149
4	16216.302	12031.992	10383.792	0	16481.211
5	21988.945	14831.589	8623.149	16481.211	0

Source: Our calculations

Also, cluster 5 displays significant differences from the other clusters, except cluster 3.

According to the distances between the centers of gravity of classes, the dividing of the set of objects in 5 clusters is justified by the dissimilarities between these above, as the clusters have distances between them, the distances are between 7198.41 and 22189.20.

Table no. 16 **Final Cluster Centers**

Variables	**Cluster**				
	1	2	3	4	5
Equity	1287.40	1824.65	9547.34	13186.00	1050.00
Net Income	172.76	522.78	1397.50	160.67	1376.00
Current Sales	2035.19	9203.18	22589.22	13022.00	23971.42
EVA	32.0269	193.3340	440.00	-793.61	932.98
Stock Price	42.12	33.49	28.94	36.58	37.45

Source: Our calculations

The cluster companies have the highest Stock Price, the lowest Current Sales, Net Income (not the lowest), the lowest EVA, and low Equity.

It is noted that the companies in cluster 5 have the highest values for all indexes except Stock Price, which is lower than with the companies in the first cluster.

A special category is represented by the companies in cluster 4, which have lower Current Sales and Equity higher than those in cluster 3. Also, the EVA is strongly negative (-793.61 million $) and Net Income has the lowest value of all clusters.

Table no.17 **ANOVA table**

	Cluster		Error		
	Mean Square	df	Mean Square	df	F
Equity	$1.855*10^8$	4	4292869.61	90	43.22
Net Income	3093115.72	4	271893.85	90	11.38
Current Sales	$9.865*10^8$	4	4589743.32	90	214.94
EVA	1412066.14	4	207581.38	90	6.80
Stock Price	478.65	4	1930.70	90	0.25

Source: Our calculations

For the identified clusters, the analyzed variables provide a great discrimination between them, as evidenced by the high values of the F statistics, defined as the intra-class variability/ inter-class variability ratio, scaled by the degrees of freedom.

Of all the variables considered in the classification process, Current Sales has the largest contribution in the classification process; the Stock Price variable having the smallest contribution. An important contribution is made by the other variables which have values greater than 1.

V. CONCLUSIONS

Given the conditions when the amount of information is very high, the Decision Making Unit is unable to identify a suitable criterion for making the best decisions regarding the activity they carry out. Also, the asymmetry existing in the collected data sets makes it impossible to describe the analyzed phenomenon using a mathematical model, especially when each objects is characterized by a large number of features. In these situations it is useful to segment the large amount of data sets, so as to enable the identification of specific typologies existing in the data sets. The clustering algorithms provide the optimal solution for handling such situations, leading finally to the identification –under conditions where the description of an evolution model is impossible –of a structure which simplifies the large amount of information and represents the basis for the future scientific efforts to be made.

Applying the classification algorithms to the data set on the 106 surveyed companies in the 4 different fields of activity – companies identified by 5 basic economic features– to analyze their activities, determines their distribution in 7 classes of companies.

As a measure of objects proximity, we used the Manhattan distance, because it is more robust than Euclidean distance. In the first stage of the scientific approach, we performed the full hierarchical classification using Ward’s method as aggregation method.

Following the application of this method to the original company sample, we observed that the original sample was structured in two classes, the first class concentrates almost all the considered companies and the second class contains atypical companies.

From the atypical company group, the Microsoft Corp. and Oracle Corp. companies are outstanding, both operating in the software field, as they are characterized by the highest values of the EVA variable, which we can consider as a global indicator of the performed activity, and in this respect such companies can be considered the most efficient.

Table no.18 Source: **Our calculations**

Variables	Cluster	
	1	2
Equity	16622.80	29840.75
Net Income	824.80	4163.50
Current Sales	41349.20	60569.25
EVA	-950.32	380.31
Stock Price	34.02	47.91

From the atypical companies we obtained a classification thereof in two subclasses: in the first subclass are grouped companies which have relatively small values of the Net Income, the lowest EVA value, but a high volume of Current Sales.

Companies in the two clusters are similar in terms of Stock Price.

In second class, we find companies which have very high Equity values, the highest Current Sales, the highest Net Income, and low EVA.

Thus, the two classes, although not containing a large number of companies, are necessary because they are patterns enabling the comparisons to other companies whose class membership is unknown.

In these classes, the companies having registered values opposite in meaning to EVA, using a large amount of Equity are identified.

Compared to the atypical companies for the majority group of companies, after summarizing the achieved results, we obtain the following representation for 5 classes, which have a high relevance for the conducted analysis, the predictor variables having high separation power:

Table no.19

Variables	Cluster				
	1	2	3	4	5
Equity	13186.00	1287.40	1824.65	9547.34	1050.00
Net Income	160.67	172.76	522.78	1397.50	1376.00
Current Sales	13022.00	2035.19	9203.18	22589.22	23971.42
EVA	-793.61	32.03	193.33	440.00	932.98
Stock Price	36.58	42.12	33.49	28.94	37.45

Source: Our calculations

After interpreting the 5 obtained classes we distinguish:

- **The first class** we identified is considered ***the class of inefficient companie**s*, characterized by destruction of value (negative EVA), the highest volume of Equity, the lowest Net Income, the largest Current Sales and medium level Stock Price.

- **The second class** is considered ***the class of less efficient companies,*** characterized by positive but low EVA, low Equity, relatively low Net Income, low Current Sales but the highest Stock Price.

- **The third class** is considered ***the class of average efficiency companies***, with EVA about $ 200 million, obtained by a higher volume of Equity, the reported Net Income of about $ 530 million, and low Stock Price, which makes them attractive to investors.

- **The fourth class** companies are considered to be ***the efficient companies***, with average EVA about $ 440 million, achieved by the highest level of Equity, the highest Net Income relative to very high Sales Current. Also, the companies in this class are characterized by the lowest level of Stock Price.

- **The fifth class** companies are ***very efficient companies***; a company in this class is characterized by the highest EVA relative to a level of Net Income similar to that achieved by an *efficient company* but the associated Equity in achieving the above is 9 times lower. The companies in this class are characterized by high Stock Price, close to that of the less efficient companies, which reports the highest price level.

The identification of the class a company belongs to, which was not employed in the determination of the class is conducted based on the following rule:

➢ ***after identifying the numeric values associated to the analyzed variables, i.e. EVA, Net Income, Current Sales, Equity, and Stock Price, one shall determine the distance toward the center of each identified class and the unclassified company is allocated to the nearest class*.**

However, from the analysis of the obtained classes, it is found that the variable which provides the highest similarity between classes is Stock Price. This fact justifies the elimination of this feature from the analysis to be conducted in future studies.

After identifying the classes and founding the properties specific to each class –as a proposition for a future study– the problem is to determine the aggregate indicators which ensure the segmentation of companies by classes using the techniques of discrimination.

***Acknowledgments*:**

″This work was co-financed from the European Social Fund through Sectorial Operational Program Human Resources Development 2007-2013, project number POSDRU/107/1.5/S/77213 „Ph.D. for a career in interdisciplinary economic research at the European standards"

BIBLIOGRAPHY:

- Altman, E.I. – "Financial Rations, Discriminant Analysis and Prediction of Corporale Bankruptcy", Journal of Finance, 1968
- Armeanu, Ş. D., Vintilă, G., Moscalu, M., Filipescu, M. O., Lazăr, P., (2012), ′Use of techniques of quantitative data analysis for the estimation of corporate bankruptcy risk′ Theoretical and applied economics, Volume XIX (2012), No. 1(566), pp. 86-102
- Bao, B.H., Bao, D. H. (1998). Usefulness of Value Added and Abnormal Economic Earnings: An Empirical Examination. Journal of Business Finance and Accounting, 25(1-2), 251-265.
- Bogeanu, A., ″ Relevance of specific tools for an entity's performance valuation. Study of case: PETROM, a joint stock company″ Zigotto Publishing House, Galaţi, 2010
- Brealey, R., Myers, S., Allen, F. (2008). Principles of Corporate Finance. McGraw Hill, New York.
- Correira, C., Flynn, D., Uliana, E., Wormald, M. (2007). Financial Management (6th edition). Cape Town: Juta & Company.
- Damodaran, A. (2001). The Dark Side of Valuation: Valuing Old Tech, New Tech, and Economy Firms, Prentice Hall, New York.
- Damadoran A. - "Value Creation and Enhancement: Back to the Future", Stern School of Business, New York, 2006

Dosch, J., Wilson, J. (2010). Process Costing and Management Accounting in Today's Business Environment. Strategic Finance, August, 36-43.

- Guatri, L., Bini, M. (2005). Nuovo trattato sulla valutazione delle aziende. Università Bocconi, Milano.
- Hall, J. M., Johnson, M. E. (2009). When should a Process be Art, Not Science? Harvard Business Review, 87(3), 58-65.
- Holler, A. (2008). Have Earning Lost Value- Relevance? Revisiting Latest Evidence on EVA. The Business Review, 10(2), 245-54.
- Lovata, M. L., Costigan, L. M. (2002). Empirical Analysis of Adopters of Economic Value Added. Management Accounting Research, 13(2), 215-228.
- MacKay, D. (2003). An Example Inference Task: Clustering. In. Information Theory, Inference and Learning Algorithms (pp. 284–292). Cambridge University Press.
- Magnanelli, B. S. (2010). Il performance management a supporto dell'allineamento dell'impresa alle esigenze del mercato: la balanced scorecard nel settore farmaceutico. In De Bernardis L. (eds), Allineare mercati, strategie e comportamenti. Dal paradigma di Bain a quello dei network, 117-138, Milan: FrancoAngeli.
- Mocciaro Li Destri, A., Dagnino, G. B. (2004). Time and Strategy: Towards a Multitemporal View of the Firm. Journal of Managerial Psychology, 18(9), 776-794.
- Mocciaro Li Destri, A., Dagnino, G. B. (2005). The Development of the Resource-Based Firm Between Value Appropriation and Value Creation. In Szulanski G., Porac J., Doz Y. (eds.) Strategy Process (Advances in Strategic Management, Vol. 22), pp.153-188, Emerald Group Publishing Limited.
- Moisello, A. M. (2003). ABC & EVA: un'integrazione possibile. La valutazione del costo di prodotto nell'ottica della creazione di valore. Economia Aziendale 2000 web, www.ea2000.it.
- Multi-varied Statistics – Paper No. 12 — Classification - SPSS http://profs.info.uaic.ro/~val/statistica/StatWork_12.pdf; 15.10.2012
- Niculescu M. - Global Strategy Diagnosis, Financial Diagnosis, Economics Publishing House, Bucharest, 2005
- Palliam, R. (2006). Further Evidence on the Information Content of Economic Value Added. Review of Accounting and Finance, 5(3), 204-215.
- Parkinson, J. (2004), 'Economic Value Added: Does it Add Value?', International Accountant, May 2004, pp. 36-8. (B)
- Picone, P. M. (2012). Conglomerate diversification strategy and corporate performance. Unpublished doctoral dissertation, University of Catania, Italy, http://hdl.handle.net/10761/1167
- Pratt, S. P., Grabowski, R.J. (2008). Cost of Capital. Applications and Examples. John Wiley & Son, Hoboken, New Jersey.
- Reed, J. (2005). Values and Corporate Responsibility: A Personal Perspective, In: Lorsch J.W., Berlowitz, L., Zelleke, A. (eds), Restoring Trust in American Business, MIT Press, Cambridge, MA.
- Singer, J. A., Millar, D.L. (2003). Value-Based Management Done Right: The EVA Implementation at Harsco. EVAluation, 5(1), 1-8.
- Stewart, G. B. (1991). The Quest for Value: A Guide for Senior Management. Harper Business, New York.
- Vâlceanu Gh., Robu V., Georgescu N. - Economic and financial analysis, "Economics" Publishing House, Bucharest, 2005
- Zubin, J. A. (1938). A technique for measuring likemindedness. Journal of Abnormal & Social Psychology (27), 508-516.
- Worthington, A. C., West, T. (2004). Australian Evidence concerning the Information Content of Economic Value-Added. Australian Journal of Management, 29(2), 201-224.
- Young, D. (1997). Economic Value Added: A Primer for European Managers. European Management Journal, 15(4), 335-343.
- http://www.evadimensions.com 25.07.2012
- http://eresearch.fidelity.com/eresearch/evaluate/fundamentals/keyStatistics.jhtml?
- http://www.evanomics.com/evastudy.shtml
- http://www.gurufocus.com/
- http://incomemagazine.ro/articles/4-ani-de-la-lehman-brothers-falimentul-care-a-schimbat-lumea-intreag
- http://www.moneyterms.co.uk/eva
- http://www.nyse.com/
- http://pages.stern.nyu.edu/~ealtman/Zscores.pdf
- http://pages.stern.nyu.edu/~adamodar/New_Home_Page/lectures/eva.html
- http://quicktake.morningstar.com/StockNet/Diagnostics.aspx?
- http://www.reuters.com/

- http://seekingalpha.com/
- http://www.stasegem.be/PRVit/PRVitList.php 31.07.2012
- www.sternstewart.com/
- www.valuebasedmanagement.net
- http://www.wikiwealth.com/
- http://www.zacks.com/
- http://www.ziare.com/articole/lehman+brothers+vinovat+criza
- http://www.wall-street.ro/articol/Interviuri/5642/Peter-Drucker-Eu-nu-cred-in-lideri.html; 17 august 2005