CLUSTERING THE COMPANIES FROM THE REAL ESTATE FIELD OF ACTIVITY FROM ROMANIA, USING STATISTICAL ANALYSIS SYSTEM AND THE R PROGRAM

Ionela-Catalina Tudorache[1]
Ana-Maria Mihaela Iordache[2]
Mihai-Tiberiu Iordache[3]

Abstract

In this article we have analyzed, from data mining point of view, the Romanian real estate market and, on the other hand, we have predicted if a new company that enters on this market would be profitable, which it means that it would belong to the big class of companies or to the non-profitable class, the small class of companies. In the analysis we used mathematical instruments such as the cluster and the discriminant analysis and the programs used for application were Statistical Analysis System and the R program. The main idea that we rely on is the fact that, after splitting the companies into two main clusters, using the cluster analysis, it is possible to calculate the discriminant function and also the discriminant scores and, further, to integrate a new company into one class using the discriminant analysis.

Key words: cluster analysis, discriminant analysis, Ward, the dendrogram, CAEN, Bucharest Stock Exchange, SAS, the R program

Choosing The Indicators And The Objects For The Analyses

The main indicators that characterized the real state of a company in the economy and that we chose in our analysis were: the fixed assets, the share capital, the receivables, the debts, the number of employees and the net income [11]. We considered the all 47 companies listed at the Bucharest Stock Exchange, these companies are in the real estate field of activity and their CAEN - National Classification of Economic Activities code was 6820. [14] The values of the indicators are expressed in Romanian RON for 2010 year, excluding the number of employees (see table 1).

Symbol	Company name	Fixed assets	Share capital	Receivables	Debts	Employees	Net income
ALIA	ALIA SA	3431154	453640	610404	1307475	14	93689
ASBU	ASTRAL SA	20933648	1560290	1144076	626323	7	683560
BAST	BASTI SA	5432214	525678	146491	609512	7	8001
CAIN	CASA ALBA - INDEPENDENTA SA	5380118	1966863	195919	17532	1	327
CHIB	CHIBRO SA	9978129	1461378	406734	321807	1	-394535

[1] Student, School of Computer Science for Business Management, Romanian - American University, Bucharest
[2] Assitant Lecturer, Phd, School of Computer Science for Business Management, Romanian -American University, Bucharest
[3] Engineer, IMSAT, Bucharest

CICO	CI-CO SA	624810	649593	529484	5821772	70	1764243
COCR	COCOR SA	233195086	12067640	8255175	86805010	32	-4170479
COKG	COMALEX SA	1670436	1424485	171776	262421	11	-288406
COET	COMPLEX COMET SA	467879	3121775	2504392	1082119	5	4336
COLI	CONCORDIA A 4 SA	1823062	2274855	247442	966098	7	234428
COBR	CORAL SA	1760456	927025	924119	1775796	16	65930
CRIO	CRIORTEX SA	5242650	1634538	504388	148892	10	47340
DONO	DORNACOM SA	437377	239743	741503	46602	8	45870
ELTG	ELECTROMINING SA	1318103	1030408	106031	503859	4	48517
GALV	GALGROS SA	10706522	10365760	1306580	550953	24	1553822
GAOY	GASTRONOM SA	5688678	766320	231590	243769	6	535226
COXT	GECSAT INTERCOM SA	582583	557737	83796	41370	3	60235
GECM	GENERALCOM SA	77574386	36678699	2075558	26811410	12	8788716
ICOA	ICOA SA	21298912	1619003	2027475	1506569	47	228257
IMSA	IMSA SA	4553898	679753	134747	2013433	37	-458603
INBI	INDUSTRIALIZAREA LAPTELUI SA	1446000	835083	11347	253239	2	289151
MACK	MACOMB SA	1108467	284715	81222	574971	1	-11582
MARD	MARCONS SA	5065625	999797	1353892	2287907	27	4481243
METV	METAV SA	112609236	49043220	1621507	11546091	33	6861635
MXTU	MIXT SA	554319	316785	13344	58244	6	155095
AOHL	NEAG GROUP SA	1252611	98750	770929	2830655	27	340853
ORZA	ORIZONT SA	2146673	371983	453862	831389	12	-1221601
PRBU	PRACTIC SA	181551876	40983171	928015	71170819	58	16751162
PRTM	PRETIM SA	275885	326748	278355	67709	1	16578
PRIB	PRIMCOM SA	23900118	3653940	328847	270441	11	1021729
RETE	RETEZATUL SA	336618	436015	36628	188036	8	2879
RLS	ROLAST S.A.	150650579	30808449	292805	120319402	9	1590626
SALT	SALTEX SA	596196	563707	11410	351607	5	-82308
SPID	SPID SA	369030	545163	178645	178094	2	-41108
STBU	STELA SA	5600882	582883	698374	1424486	3	51202
TCII	TCI INVEST SA	21647171	1207500	349866	212095	19	17981
TXIN	TEXIND SA	24724299	685085	219512	754339	11	-86254
AVIA	TRANSILVANIA	4968089	3921100	264068	2170682	7	155193

	INVEST H.I. SA						
MRFA	TRANSPORT AUTO MARFA SA	1788496	848338	448907	172000	9	520215
TUSB	TURDAPAN SA	280399	884155	438112	517951	6	3276
SCDM	UNIREA SHOPPING CENTER SA	520987582	7443720	15891598	108635944	170	5989947
UNIT	UNITEH SA	11328290	1090900	185546	68085	4	-207003
URBC	URBAN SA	1925520	659970	86365	23432	2	-138224
UTGR	UTILAJ GREU SA	3291291	1691468	1794055	549468	22	308785
VIRU	VIDRARU SA	11876062	716000	313263	938696	11	969017
VINA	VINALCOOL SA	7823723	1184565	4552301	1522017	1	203442
VIRO	VIROLA - INDEPENDENTA SA	917576	346480	7043	61574	1	1508

Table 1. The listed-companies from the real estate market

The indicators chosen in analysis were taken from the balance sheet and other financial reports. They have some particularities, for examples the share capital refers to the amount of money that belong to shareholders, the total receivables represents the money that will come in the company, the total debts includes the short-term and also the long-term debts, the net income could be profit or loss and the employees represents the number of permanent employees of a company. [15]

The Cluster Analysis

The cluster analysis involves using a method like the hierarchical or the partitioning algorithm in order to split the observations, all 47 companies from real economy, into clusters. The most important thing we consider here is that the variance between classes should be maxim, while the variance within classes must have a minimum value. This rule is also known as the homogeneity of classes.

Due to the fact that the units of measurement for our variables are different, RON for monetary variables and units for employees, it was necessary to standardize the data matrix. The standardization is the process in which we obtain new values for the variables starting from the old value minus the mean and all divide by standard deviation.

But, before standardizing the matrix of values of indicators [1] there are two problems that occur: the outlier problem and the choice of the method we use. An outlier is an observation that has the highest values for the variables we take into account and that can influence the mean that much, that this one would become unrepresentative for the sample we analyze.

In order to have the results as better as possible, it is necessary to first analyze the observations we have. After a closer look, we noticed that the observation number 41, that represents the Unirea Shopping Center SA, is an outlier. If we want to validate the obtained results, it is necessary to eliminate the outliers. Therefore, from this point on, we will not take into account the observation number 41.

To split the observations into two classes named big companies and small companies we have to choose between two alternatives: the hierarchical methods or the partitioning algorithm.

We used the Ward method, as one from five hierarchical methods and the k-means algorithm as the partitioning algorithm to split the observations into two classes, named the clusters.

In SAS 9.2 we used the following procedures [12]:

```
ods graphics on;
proc cluster data=work outtree=tree method=WARD plots=all;
run;
proc tree;
RUN;
proc fastclus data=work out=w1 maxclusters=2;
run;
```

The first procedure is for the Ward method, using the imported data named *work*, the second is for building the dendrogram, while the last one is used for the k-means algorithm. The last procedure has determined already the number of clusters, maxclusters=2 [13].

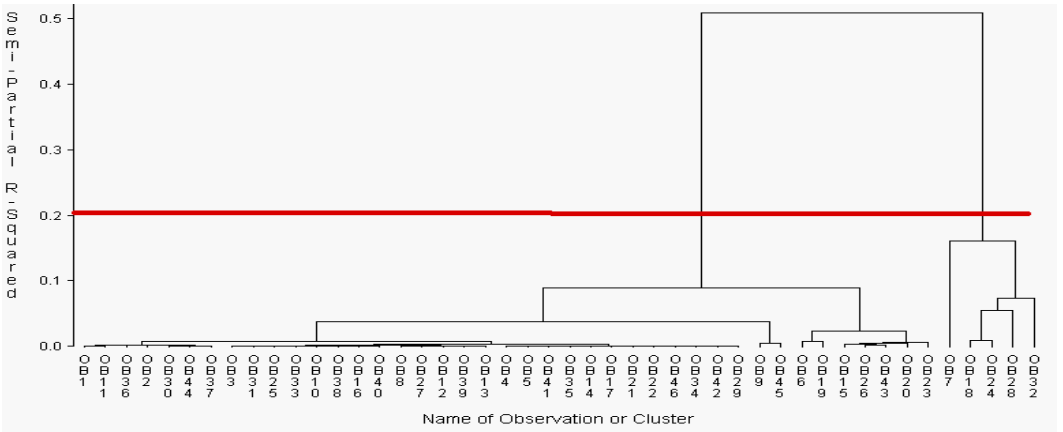

Figure 1. The dendrogram resulted from the Ward method

Figure 1 shows the way in which the observations are merging into clusters and the distance at which the classes are formed, from leafs - 46 classes, to the top of the tree - 1 class.

The red thick line represents the cut in the dendrogram. We decided to take into account two classes: the big class of companies, composed by the companies that has the symbols: COCR, GECM, METV, PRBU and RLS, and the small class of companies composed by the rest of them which are not in the first class.

We considered a company to be big if the values for the variables are high, no matter if we took into account the debts or if the net income is negative (loss) for that period of time (2010).

NAME	PARENT	NCL	NAME	PARENT	NCL	NAME	PARENT	NCL
OB22	CL45	46	OB11	CL31	46	CL18	CL16	18
OB46	CL45	46	CL34	CL30	34	OB15	CL15	46
OB34	CL44	46	OB41	CL30	46	CL17	CL15	17
OB42	CL44	46	CL36	CL29	36	CL19	CL14	19
OB14	CL43	46	OB13	CL29	46	CL22	CL14	22
OB17	CL43	46	CL38	CL28	38	OB9	CL13	46
OB25	CL42	46	CL32	CL28	32	OB45	CL13	46
OB33	CL42	46	OB30	CL27	46	CL15	CL12	15
CL45	CL41	45	OB44	CL27	46	OB20	CL12	46
CL44	CL41	44	OB8	CL26	46	CL12	CL11	12
OB3	CL40	46	OB27	CL26	46	OB23	CL11	46
OB31	CL40	46	CL30	CL25	30	OB6	CL10	46
CL43	CL39	43	OB35	CL25	46	OB19	CL10	46
OB21	CL39	46	CL35	CL24	35	CL16	CL9	16
OB10	CL38	46	CL28	CL24	28	CL14	CL9	14
OB38	CL38	46	CL27	CL23	27	OB18	CL8	46
CL41	CL37	41	OB37	CL23	46	OB24	CL8	46
OB29	CL37	46	CL25	CL22	25	CL10	CL7	10
OB12	CL36	46	CL33	CL22	33	CL11	CL7	11
OB39	CL36	46	CL26	CL21	26	CL9	CL6	9
CL40	CL35	40	CL29	CL21	29	CL13	CL6	13
CL42	CL35	42	CL31	CL20	31	CL8	CL5	8
OB4	CL34	46	OB36	CL20	46	OB28	CL5	46
OB5	CL34	46	CL24	CL19	24	CL5	CL4	5
CL39	CL33	39	CL21	CL19	21	OB32	CL4	46
CL37	CL33	37	OB2	CL18	46	CL6	CL3	6
OB16	CL32	46	CL23	CL18	23	CL7	CL3	7
OB40	CL32	46	OB26	CL17	46	OB7	CL2	46
OB1	CL31	46	OB43	CL17	46	CL4	CL2	4
			CL20	CL16	20	CL3	CL1	3
						CL2	CL1	2
						CL1		1

Table 2. The output from cluster procedure - Ward

In Table 2 it is presented the output of the first SAS procedure, named *tree*. It is shown here the way in which the main big class is formed, starting from leafs and ending at the top of the tree.

Symbol	Cluster	Symbol	Cluster	Symbol	Cluster	Symbol	Cluster	Symbol	Cluster	Symbol	Cluster
ALIA	1	COET	1	COXT	1	MXTU	1	SALT	1	UNIT	1
ASBU	1	COLI	1	GECM	2	AOHL	1	SPID	1	URBC	1
BAST	1	COBR	1	ICOA	1	ORZA	1	STBU	1	UTGR	1
CAIN	1	CRIO	1	IMSA	1	PRBU	2	TCII	1	VIRU	1
CHIB	1	DONO	1	INBI	1	PRTM	1	TXIN	1	VINA	1

CICO	2	ELTG	1	MACK	1	PRIB	1	AVIA	1	VIRO	1
COCR	2	GALV	1	MARD	1	RETE	1	MRFA	1		
COKG	1	GAOY	1	METV	2	RLS	2	TUSB	1		

Table 3. The output of applying the k-means algorithm

Table 3 shows the output of the k-means algorithm, named *w1* in the third SAS procedure. Now, the big class of companies is formed from: CICO, COCR, GECM, METV, PRBU and RLS.

The difference between the two splitting methods is that the company which symbol is CICO is a small company, according to the Ward method and a big company from the k-means algorithm analysis. For the discriminate analysis, we will consider the clusters resulted from the k-means algorithm, presented in Table 3.

The Discriminant Analysis

The discriminant analysis is based on the theory that starting from a learning sample, the observations that we know their affiliation to a class, we can determine the affiliation of another observation, that does not belong to the learning sample to one of the clusters. [2] To determine the discriminant function, it is necessary to calculate the eigenvector that corresponds to the highest eigenvalue of the matrix calculated by the formula $\Sigma_w^{-1} \cdot \Sigma_b$[7], where:

- Σ_w^{-1} is the inverse of the covariance matrix that measures the variance within classes. This matrix is calculating like: $\Sigma_w = \frac{T_1}{T}\Sigma_{w_1} + \frac{T_2}{T}\Sigma_{w_2}$, where:
  - T_1 is the number of observations, objects in class 1, there are 40 in our case
  - T_2 is the number of observations, objects in class 2, 6 in this case
  - T is the total number of observations, there are 46 observations in our study
  - Σ_{w_1} is the covariance matrix that corresponds to class 1
  - Σ_{w_2} is the covariance matrix that corresponds to class 2
- Σ_b is the covariance matrix that measures the veriability between clases. We can calculate it like: $\Sigma_b = \Sigma - \Sigma_w$, where:
  - Σ is the covariance matrix of the entire sample, all 46 observations.

Considering this, we will present the steps of calculated the final matrix, that are presented in the next figures. In this way, in figure 2 we calculated the Σ matrix, figure 3 presented the Σ_{w_1} matrix and so on.

sigma	*Fixed_assets*	*Share_capital*	*Receivables*	*Debts*	*Employees*	*Net_income*
Fixed_assets	2.40836E+15					
Share_capital	4.10832E+14	1.22334E+14				
Receivables	4.0078E+13	3.55778E+12	1.92849E+12			
Debts	1.02817E+15	1.64406E+14	1.3736E+13	5.5062E+14		
Employees	304642548.9	61163795.7	5489106.751	104834035.7	228.5538752	
Net_income	6.47023E+13	2.55436E+13	-2.25067E+11	2.28041E+13	20616198.06	9.12623E+12

Figure 2. The matrix Σ

sigmaw1	*Fixed_assets*	*Share_capital*	*Receivables*	*Debts*	*Employees*	*Net_income*
Fixed_assets	4.9826E+13					
Share_capital	3.1664E+12	2.85129E+12				
Receivables	9.71649E+11	3.35068E+11	7.26479E+11			
Debts	1.83976E+11	48619737372	2.39485E+11	5.14241E+11		
Employees	23640875.03	3484397.396	2054173.89	4105796.049	103.6775	
Net_income	9.20599E+11	3.87576E+11	1.47839E+11	1.73646E+11	2423125.096	6.4792E+11

Figure 3. The matrix Σ_{w_1}

sigmaw2	*Fixed_assets*	*Share_capital*	*Receivables*	*Debts*	*Employees*	*Net_income*
Fixed_assets	2.91094E+15					
Share_capital	-4.72485E+14	1.54684E+14				
Receivables	9.81208E+13	-2.88907E+13	8.26524E+12			
Debts	1.36372E+15	-3.03762E+14	1.15964E+13	1.57511E+15		
Employees	446873339.3	43019426.36	3931602.8	-85659663.72	310.96	
Net_income	-1.48755E+14	6.65259E+13	-1.35236E+13	-1.1044E+14	65092488.2	4.93963E+13

Figure 4. The matrix Σ_{w_2}

sigma w	Fixed_assets	Share_capital	Receivables	Debts	Employees	Net_income
Fixed_assets	4.23014E+14	-5.88751E+13	1.36433E+13	1.78036E+14	78845109.5	-1.86023E+13
Share_capital	-5.88751E+13	2.26555E+13	-3.47699E+12	-3.95789E+13	8641140.305	9.01431E+12
Receivables	1.36433E+13	-3.47699E+12	1.7098E+12	1.72082E+12	2299055.922	-1.63539E+12
Debts	1.78036E+14	-3.95789E+13	1.72082E+12	2.05896E+14	-7602742.182	-1.42542E+13
Employees	78845109.5	8641140.305	2299055.922	-7602742.182	130.7143478	10597389.85
Net_income	-1.86023E+13	9.01431E+12	-1.63539E+12	-1.42542E+13	10597389.85	7.00641E+12

Figure 5. The matrix Σ_w

sigma w at -1	Fixed_assets	Share_capital	Receivables	Debts	Employees	Net_income
Fixed_assets	7.32729E-15	7.80045E-15	-3.27236E-14	-4.80858E-15	-4.54848E-09	-1.12287E-15
Share_capital	7.80045E-15	1.60412E-13	1.6151E-13	1.5245E-14	-8.86883E-09	-1.03545E-13
Receivables	-3.27236E-14	1.6151E-13	1.3069E-12	5.90883E-14	-2.40203E-08	1.66914E-13
Debts	-4.80858E-15	1.5245E-14	5.90883E-14	1.17312E-14	1.26268E-09	3.36791E-15
Employees	-4.54848E-09	-8.86883E-09	-2.40203E-08	1.26268E-09	0.013422236	-2.40052E-08
Net_income	-1.12287E-15	-1.03545E-13	1.66914E-13	3.36791E-15	-2.40052E-08	3.55085E-13

Figure 6. The matrix Σ_w^{-1}

sigma b	Fixed_assets	Share_capital	Receivables	Debts	Employees	Net_income
Fixed_assets	1.98534E+15	4.69707E+14	2.64348E+13	8.5013E+14	225797439.4	8.33046E+13
Share_capital	4.69707E+14	9.96783E+13	7.03477E+12	2.03985E+14	52522655.39	1.65293E+13
Receivables	2.64348E+13	7.03477E+12	2.18697E+11	1.20152E+13	3190050.83	1.41032E+12
Debts	8.5013E+14	2.03985E+14	1.20152E+13	3.44724E+14	112436777.9	3.70583E+13
Employees	225797439.4	52522655.39	3190050.83	112436777.9	97.83952741	10018808.21
Net_income	8.33046E+13	1.65293E+13	1.41032E+12	3.70583E+13	10018808.21	2.11983E+12

Figure 7. The matrix Σ_b

12.13757608	2.750677201	0.167543368	5.216481461	9.62862E-07	0.467034006
97.43479842	21.72214563	1.378837527	41.71463473	1.05108E-05	3.785703337

104.1563695	23.47280263	1.425697539	44.68304584	1.1229E-05	4.089666249
9.714717618	2.191621826	0.142784755	4.042604861	1.37974E-06	0.38927478
-11726615.31	-2623712.182	-163747.4298	-4909682.24	-0.35478272	-429000.721
-19.42976471	-4.378910457	-0.256923369	-8.44994034	-3.572E-06	-0.93264465

Figure 8. The calculated matrix: $\Sigma_w^{-1} \cdot \Sigma_b$

Figures from 2 to 8 show the steps for calculating the matrix we need, presented in Figure 8. For this matrix we have to determine the highest eigenvalue and the eigen vector that is associated to the highest eigenvalue.

Using the program R, after importing the matrix from Figure 8 from a text file (that is tab-delimited), we determined the eigenvalues and the eigenvectors, presented in figure 9.

```
> a<-read.table("date_R.txt",sep="\t")
> e<-eigen(a)
> e
$values
[1] 37.66250458  0.86149717 -0.16351429 -0.15119912 -0.13884608 -0.02984552

$vectors
              [,1]          [,2]          [,3]          [,4]          [,5]
[1,] -1.028226e-06 -3.098016e-07 -3.114837e-07 -3.507564e-06 -2.307102e-06
[2,] -8.105875e-06  4.423382e-07  4.105311e-05  2.002862e-06  1.631992e-05
[3,] -8.708495e-06  5.270620e-07 -2.606821e-05 -1.288556e-05  3.015101e-06
[4,] -8.031893e-07  3.738412e-07 -2.458066e-05  6.694788e-06  7.897788e-07
[5,]  1.000000e+00  1.000000e+00  1.000000e+00 -1.000000e+00  1.000000e+00
[6,]  1.578577e-06 -1.551676e-06  4.825559e-05  1.240371e-05 -4.743931e-05
              [,6]
[1,] -3.388565e-06
[2,]  1.798299e-06
[3,]  7.464420e-05
[4,]  3.345374e-06
[5,] -1.000000e+00
[6,]  1.560742e-05
```

Figure 9. The R output

Using the first calculated eigenvector, we determinated the discriminant function after applying the formula [8]:

$D(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots$, where $\beta_0 = -[\beta_1 \overline{X_1} + \beta_2 \overline{X_2} + \cdots]$

We mention here that $\{X_1, X_2 \ldots X_n\}$ represents the variables we took into account, $\{\overline{X_1}, \overline{X_2}, \ldots \overline{X_n}\}$ is the mean for these variables, and β is the used eigenvector.
Therefore, the discriminant function is:

$$D(x) = 24.51664 - 0.000001028226X_1 - 0.0000008105875X_2 - 0.000008708495X_3 - 0.0000008031893X_4 + X_5 + 0.000001578577X_6$$

The discriminant scores represent a number calculated for each observation using the discriminant function. If this number is higher than zero, the object belongs to class 1, and if it is smaller than zero, the observation belongs to class 2.

Symbol	Class	Scores	Symbol	Class	Scores	Symbol	Class	Scores
ALIA	1	28.40297708	ICOA	1	29.79821513	AVIA	1	19.43181632
ASBU	1	-0.659808121	IMSA	1	52.76866139	MRFA	1	27.7637648
BAST	1	23.75235741	INBI	1	24.50715889	TUSB	1	25.28551258
CAIN	1	16.67063184	MACK	1	22.95869254	UNIT	1	13.98705461
CHIB	1	9.648985331	MARD	1	38.94362128	URBC	1	23.01268867
COKG	1	30.48243785	MXTU	1	29.77174368	UTGR	1	26.18396985
COET	1	3.833304513	AOHL	1	41.6995154	VIRU	1	20.77266028

COLI	1	25.23741544	ORZA	1	27.45924985	VINA	1	-24.03310983
COBR	1	28.5851566	PRTM	1	22.51585197	VIRO	1	24.18391169
CRIO	1	23.36376445	PRIB	1	6.511987868	CICO	2	86.84565039
DONO	1	25.45019741	RETE	1	31.35164234	COCR	2	-341.2369077
ELTG	1	25.07462817	SALT	1	27.93499006	GECM	2	-98.71460614
GALV	1	19.73753151	SPID	1	23.93163558	METV	2	-110.58788
GAOY	1	22.67853497	STBU	1	14.14010629	PRBU	2	-176.182412
COXT	1	25.79764753	TCII	1	17.09090575	RLS	2	-243.0373589
			TXIN	1	6.885505412			

Figure 10. The discriminant scores for the observations

In Figure 10, we calculated, after applying the discriminant function, the scores for each object from the studied observations. After studying the matrix from figure 10 it can be observed a different affiliation to classes for the companies with symbol VINA, CICO and ASBU.

Conclusions

Let's consider that we have a new company, which is not used in the presented study, because it is not now listed at the Bucharest Stock Exchange, and it has the same CAEN code, the domain of activity, and about that we do not know the affiliation to a class.Let his be the company Santer System S.A., symbol FIRA - it was a company delisted from stock trading, with the following information for 2010: the fixed assets: 21873479 RON; the share capital: 2996500 RON; the receivables: 10940287 RON; the debts: 9690329 RON; the net income: - 99572 RON; the number of employees: 9. The value of the score, as it is calculated in the study, is: -9.46E+01. This score is negative, which means that this company belongs to class 2, being a big company, with high indicators.

In conclusion, we can mention that this type of analysis can be made for each market, taking into account the possible study of all firms.

Bibliography

1. Anghelache Constantin, Tratat de statistică teoretică și economică, Editura Economică, București, 2008, ISBN 9789737093806
2. Fernandez George, Discriminant Analysis, A Powerful Classification Technique in Data Mining, Statistics and Data Analysis, http://www2.sas.com/ proceedings/ sugi27/ p247-27.pdf
3. Hallahan Charles, Atkinson Linda, Introduction to SAS Enterprise Guide 4.1 for Statistical Analysis, http://www2.sas.com/proceedings/sugi31/109-31.pdf
4. Leslie R. Scott, Building Reports and Documents in SAS Enterprise Guide 4.2: How to Create a Single Document from Multiple Reports and Tasks, http:// www.wuss.org/proceedings10
5. Meyers L., Gamst G., Guarino A.J., Data analysis using SAS Enterprise Guide, Cambridge University Press, 2009, ISBN: 9780511601842
6. Plemmons Howard, Working with a DBMS using SAS Enterprise Guide, SAS Global Forum, Las Vegas, Nevada, 2011
7. Ruxanda Gheorghe, Analiza multidimensionala a datelor – curs Scoala Doctorala, Bucuresti, ASE, București, 2009

8. Ruxanda Gheorghe, Analiza datelor, Editura ASE, Bucuresti, 2001, ISBN 9735940272
9. Sarma Kattamuri, Using SAS Enterprise Miner for Forecasting, http:// www.sasenterpriseminer.com /documents/SAS%20EM%20for%20Forecasting.pdf
10. SAS Enterprise Guide, Delivering the power of SAS Analytics with easy deployment of results in a Windows client interface, http://www.sas.com/resources/ factsheet /101431_0311.pdf
11. Văcărel Iulian, Bistriceanu Gh, Bercea Florian, Anghelache Gabriela, Bodnar Maria, Moşteanu Tatiana, Georgescu Florin, Finanţe publice, ediţia a IV-a, Editura Didactică şi pedagogică, Bucureşti, 2004, ISBN: 973-30-2635-2
12. http://www.sas.com/technologies/analytics/
13. http://support.sas.com/documentation/onlinedoc/guide
14. www.bvb.ro
15. www.mfinance.ro