

# Usefulness of K-means Method in Detection Corporate Crisis

*Joanna DYCZKOWSKA*<sup>\*</sup>

## Introduction

Information abundance, its accessibility but also existing information noise, caused that science about classification principles – numerical taxonomy – has been developed in response to the need of finding order in a great number of objects characterized by a wide variety of features. Taxonomic methods are applied in many scientific disciplines, including economy. In the latter case they are used especially in comparisons of:

- companies – respecting financial situation (Sojak, Stawicki, 2001),
- cities, communities, countries – respecting economic and social development (Becla, Zielińska, 2003, p. 139).

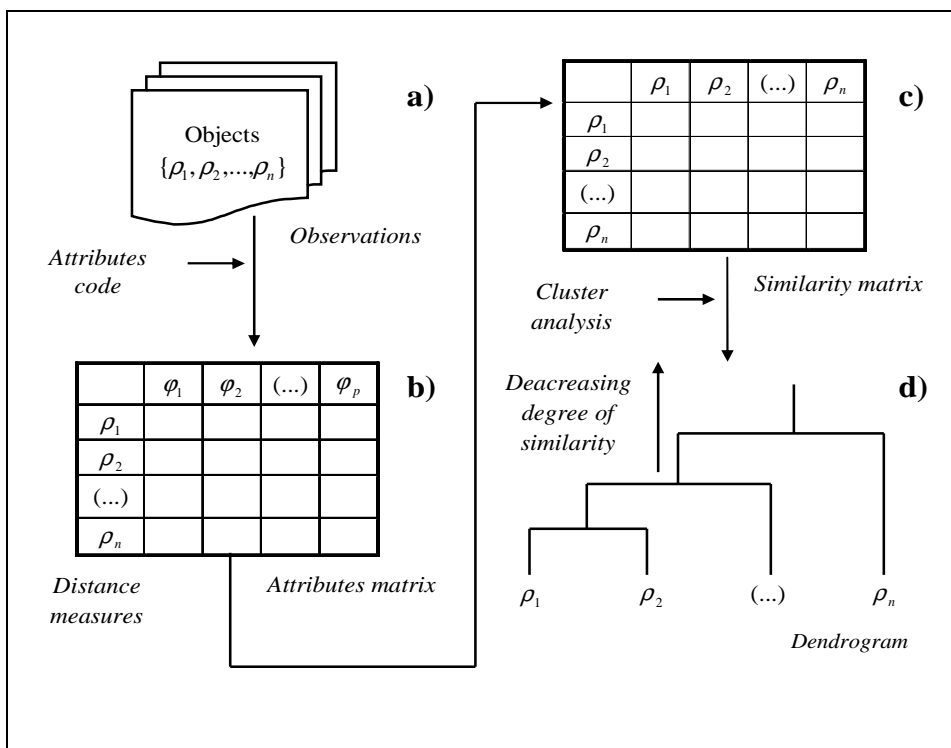
Numerical taxonomy contains different methods of which cluster analysis is particularly popular. As Aldenderfer and Blashfield (1984) define "[...] clustering method is a multivariate statistical procedure that starts with a data set containing information about a sample of entities and attempts to reorganize these entities into relatively homogeneous groups."

The aim of this paper is to prove effectiveness of using k-means clustering method in detecting corporate crisis. Classification with this method will contribute to determine position of distressed companies in a whole examined population. Each object which is assigned to a specific cluster is similar to other objects building up this group. Hence it is possible to determine levels of common attributes for each cluster. Division of the examined sample into classes will indicate differences between objects belonging to separate groups, as well.

---

<sup>\*</sup> Dr Joanna Dyczkowska – assistant professor; Department of Controlling, Faculty of Management, Computer Science and Finance, Wrocław University of Economics, ul. Komandorska 118-120, 53-345 Wrocław, Republic of Poland; <joanna.dyczkowska@gmail.com>.

**Fig. 1: Steps in numerical taxonomy**



Source: Compare with: (Silva, Campbell, Eastaugh, Buxton, 2008, p. 177-186)

To demonstrate usefulness of the aforementioned method a set of construction companies quoted at Warsaw Stock Exchange in 2003 was selected. This decision is justified by the fact that financial statements of such companies are easily available to researchers and investors.

Following managerial announcements, it was identified that nine publicly quoted building contractors went bankrupt or made an agreement with creditors in 2004-2005. This includes:

- in 2004: *Instal Lublin* (bankruptcy declaration with the possibility of concluding an arrangement with creditors), *Energomontaż Pólnoc* (arrangement with creditors), *Elkop* (bankruptcy declaration with the possibility of concluding an arrangement with creditors), *Energoaparatura* (bankruptcy declaration with the possibility of concluding an arrangement with creditors), *Naftobudowa* (arrangement with creditors),

- in 2005: *Elektromontaż Warszawa* (bankruptcy declaration with the possibility of concluding an arrangement with creditors), *Pekabex* (bankruptcy declaration including liquidation of the bankrupt), *Bick* (bankruptcy declaration including liquidation of the bankrupt), *Elektromontaż Export* (arrangement with creditors).

Results of conducted analysis should convince investors to use k-means method as a tool supporting strategic decisions concerning their financial portfolio. The only problem, that seems to be relevant, is time needed to collect complete and comparable financial information for such a number of companies.

## Selection of diagnostic variables

There exist two basic concepts in taxonomy: object and dimension. The object of classification stands for finished set of elements  $P$  (i.e construction companies), whereas dimension stands for attributes  $\Phi$  (i.e diagnostic ratios reflecting financial situations of examined companies).

$$P = \{\rho_1, \rho_2, \dots, \rho_n\} \quad (1)$$

$$\Phi = \{\varphi_1, \varphi_2, \dots, \varphi_p\}$$

where  $P$  = set of examined companies,  
 $\rho_i$  = examined company ( $i = 1, 2, \dots, n$ ),  
 $\Phi$  = set of attributes,  
 $\varphi_j$  = attribute ( $j = 1, 2, \dots, p$ ).

Set of attributes  $\Phi$  maps set of companies  $P$  into real numbers, what is expressed in the following way [13]:

$$\varphi_j : P \rightarrow \xi \subset R \quad (2)$$

where  $\xi$  = map,  
 $R$  = set of real numbers.

Multitude of attributes is a sign of multidimensional space which contain examined objects. Therefore all observations concerning particular companies are expressed in a form of matrix in the way that for each  $P_i \in R$  can be assigned  $X_{ij} \in \xi$ :

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (3)$$

where  $x_{ij}$  = volume of  $j^{th}$  variable in  $i^{th}$  company.

**Tab. 1: Data matrix**

No	Company	$X_1$	$X_2$ [days]	$X_3$ [%]	$X_4$ [days]	$X_5$	$X_6$ [%]	$X_7$
1.	Bick	0.33	724	-225.0	1846	0.27	-60.7	1.41
2.	Budimex	0.88	113	0.1	174	0.63	0.1	0.32
3.	Budopol	0.72	118	1.8	180	2.65	4.9	0.93
4.	Echo	2.46	171	3.3	305	0.41	1.4	0.45
5.	Elektrobudowa	1.40	148	1.3	126	1.47	2.0	0.51
6.	Elektromontaż Warszawa	0.60	159	-23.5	251	0.82	-19.2	0.55
7.	Elektromontaż Export	0.38	276	-77.0	657	0.57	-43.6	1.12
8.	Elkop	0.59	96	-35.9	219	1.21	-43.3	0.80
9.	Energoparatura	1.05	107	3.0	130	1.64	4.9	0.73
10.	Energomontaż Południe	1.71	84	0.5	97	1.22	0.6	0.42
11.	Energomontaż Północ	0.94	205	-32.8	273	0.91	-30.0	0.61
12.	Energopol	1.30	119	0.2	145	1.23	0.3	0.41
13.	Hydrobudowa Śląsk	1.10	69	0.5	86	1.74	0.9	0.64
14.	Instal Kraków	2.15	87	0.7	74	1.65	1.1	0.33
15.	Instal Lublin	1.22	222	-25.0	198	1.18	-29.6	0.63
16.	Instal Export	1.25	364	0.2	402	0.53	1.1	0.60
17.	Mitex	1.76	140	1.9	124	1.38	2.7	0.68
18.	Mostostal Płock	4.48	102	-5.6	58	1.14	-5.2	0.18
19.	Mostostal Warszawa	1.12	112	-15.6	157	0.97	-15.1	0.40
20.	Mostostal Export	1.17	389	-68.2	737	0.24	-16.7	0.54
21.	Naftobudowa	1.22	131	8.2	175	1.51	12.4	0.64
22.	Pekabex	0.51	142	-34.3	310	0.98	-33.7	0.93

No	Company	$X_1$	$X_2$ [days]	$X_3$ [%]	$X_4$ [days]	$X_5$	$X_6$ [%]	$X_7$
23.	Pemug	0.60	170	3.7	351	1.17	4.3	1.24
24.	Polnord	0.95	148	0.2	180	1.04	0.2	0.72
25.	Prochem	1.73	113	1.6	108	1.20	1.9	0.27
26.	Projprzem	2.39	62	3.5	49	1.81	6.4	0.28
27.	Warbud	1.30	96	1.2	82	2.86	3.4	0.69
<i>min <math>x_j</math></i>		0.33	62	-225.0	49	0.24	-60.7	0.18
<i>max <math>x_j</math></i>		4.48	724	8.2	1846	2.86	12.4	1.41
<b>Range / <math>R_j</math></b>		4.15	661	233.1	1797	2.62	73.1	1.23

Source: Own presentation

In order to classify companies using k-means method a set of seven diagnostic variables has been selected. Three criteria have been taken into account while choosing ratios: content-related, formal and statistical.

Respecting content-related criteria, selected variables should describe examined problem in a relevant and complete way. It means that chosen attributes should unambiguously signal forthcoming crisis, therefore the four tentative criteria of assessment financial situation were taken into account: financial liquidity, profitability, financial leverage and effectiveness. Set of seven diagnostic variables representing the aforementioned criteria was computed for twenty seven construction companies quoted at Warsaw Stock Exchange (See Table 1).

In order to meet formal requirement, only those objects were selected in the next step for which it was possible to gather complete financial information needed to calculate ratios. The examination excluded also these companies that declared bankruptcy in 2003, because content of their financial statements could distort results of the cluster analysis.

Attributes characterizing financial situation are expected to meet statistical criteria as well, what became the next step in a whole research procedure. Correctness of variables selection was examined using two coefficients. The first one – coefficient of variation – measures a dispersion of particular financial ratios calculated for each company and indicates which attribute provides significant discriminant power. Diagnostic variables which are selected for classification purpose are

required to be diversified within a group of examined objects. Coefficient of variation is expressed by the following formula:

$$V_j = \frac{s_j}{\bar{x}_j} \quad (4)$$

where  $V_j$  = coefficient of variation,  
 $s_j$  = standard deviation of  $j^{th}$  variable,  
 $\bar{x}_j$  = arithmetic mean of  $j^{th}$  variable.

It was assumed that financial ratios would be excluded from research if coefficient of variation reached a level lower or equal to 10%. (Compare with: Nowak, 1997, p. 12; Pawłowicz, 2001, p. 64-66; Witkowska, 2002, p. 72-73). Table 2 presents results of calculation and proves that the condition was met because each variable could be characterized as diversified within the examined group of companies.

**Tab. 2: Coefficients of variation for standardized ratios<sup>1</sup>**

Diagnostic ratios	Coefficient of variation
$X_1$ – Current ratio (current assets/current liabilities)	86.82%
$X_2$ – Average collection period (accounts receivables·365 days/net sales revenues)	123.47%
$X_3$ – Return on sales (net profit or loss/net sales revenues)	22.65%
$X_4$ – Liabilities payment time (current liabilities·365 days/net sales revenues)	154.95%
$X_5$ – Assets turnover (net sales revenues/total assets)	64.99%
$X_6$ – Return on assets (net profit or loss/total assets)	36.67%
$X_7$ – Debt ratio (total liabilities/total assets)	65.46%

Source: Own presentation

Pearson's correlation coefficient is the second measure that helps to evaluate a level of resemblance between variables. The aforementioned coefficient was computed for each pair of variables, what was presented in the following matrix:

---

<sup>1</sup>To calculate coefficients of variation, standardized values of particular ratios were used.

**Tab. 3: Matrix of Pearson's correlation coefficients**

<b>Diagnostic ratios</b>	<b><math>X_1</math></b>	<b><math>X_2</math></b>	<b><math>X_3</math></b>	<b><math>X_4</math></b>	<b><math>X_5</math></b>	<b><math>X_6</math></b>	<b><math>X_7</math></b>
X1 – Current ratio	1.00	–0.32	0.36	–0.39	<b>0.11</b>	0.42	–0.68
X2 – Average collection period	–0.32	1.00	–0.88	0.95	–0.58	–0.60	0.55
X3 – Return on sales	0.36	–0.88	1.00	–0.95	0.48	0.82	–0.59
X4 – Liabilities payment time	–0.39	0.95	–0.95	1.00	–0.54	–0.68	0.65
X5 – Assets turnover ratio	<b>0.11</b>	–0.58	0.48	–0.54	1.00	0.49	–0.08
X6 – Return on assets	0.42	–0.60	0.82	–0.68	0.49	1.00	–0.51
<b>X7 – Debt ratio</b>	–0.68	0.55	–0.59	0.65	–0.08	–0.51	1.00

Source: Own presentation

Interpretation of the correlation matrix allow to select diagnostic variables between which there existed weak linear relation. This proceeding required to fix a threshold value for the Pearson's correlation coefficient, marked as  $r^*$ , based on the following formula:

$$r^* = \left[ \frac{t_\alpha^2}{t_\alpha^2 + n - 2} \right]^{\frac{1}{2}} \quad (5)$$

where  $t_\alpha^2$  = is read from the table of Student's t distribution for  $n-2$  degrees of freedom and assumed level of significance  $\alpha = 0.05$ .

Exceeding threshold value  $r^* = 0.276$  was the reason for removing five variables from the examination (Table 3). Finally, the classification with k-means method was based on two financial variables: current ratio and assets turnover ratio between which there existed weak positive linear relation. The Table 3 indicates also that between the fifth and the seventh variable there was a weak correlation. Debt ratio, however, was excluded from the research because of its little diversion within the examined group of companies in comparison to the other two selected ratios.

## Cluster analysis using k-means method

Elements of set P are compared with each other using resemblance measures including similarity or dissimilarity coefficients. Since the purpose of this research is to recognize resemblance concerning financial situation of construction companies, the following step is connected with computation of a dissimilarity coefficient called the Euclidean metric, defined as a distance between two objects when they are perceived as points in the two-dimensional space created by their attributes. The aforementioned coefficient is expressed by the following formula:

$$d(\rho_i, \rho_j) = d_{ij} = \left[ \sum_{p=1}^P (z_{ip} - z_{jp})^2 \right]^{\frac{1}{2}} \quad (6)$$

It can be assumed that the greater distance between examined objects, the less similarity between their financial conditions. However, in order to calculate matrix of Euclidean distance another problem should be taken into consideration. Each diagnostic variable is characterized by a unit of measure and an order of magnitude. Calculation of Euclidean distance requires convert original attributes into unit- and dimensionless ones. This procedure is called standardization process of data matrix. A decision which standardizing function should be applied depends on which taxonomic method and resemblance measure are going to be used (See: Kukuła, 2000 p. 82-104; Nowak, 1997, p. 12; Romesburg, 2004). In this case the following standardizing function was applied:

$$z_{ij} = \frac{x_{ij} - \min x_j}{R_j} \quad (7)$$

where  $z_{ij}$  = standardized  $j^{\text{th}}$  variable in  $i^{\text{th}}$  object,  
 $\min x_j$  = minimum quantity of  $j^{\text{th}}$  variable,  
 $R_j$  = range of  $j^{\text{th}}$  variable.

A range of the standardized variables includes values from 0 to 1. Table 4 presents results of the computations. It is visible that each column of the standardized data matrix contains  $z_{ij} = 0$ , when  $x_{ij} = \min x_j$ , and  $z_{ij} = 1$  when  $x_{ij} = \max x_j$ .



**Tab. 4: Standardized data matrix**

No	Company	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>
1.	Bick	0.0000	1.0000	0.0000	1.0000	0.0097	0.0000	1.0000
2.	Budimex	0.1330	0.0761	0.9653	0.0694	0.1470	0.8316	0.1140
3.	Budopol	0.0941	0.0840	0.9728	0.0732	0.9184	0.8974	0.6087
4.	Echo	0.5130	0.1635	0.9791	0.1426	0.0650	0.8496	0.2208
5.	Elektrobudowa	0.2588	0.1287	0.9707	0.0428	0.4677	0.8578	0.2700
6.	Elektromontaż Warszawa	0.0650	0.1454	0.8640	0.1126	0.2184	0.5681	0.3040
7.	Elektromontaż Export	0.0115	0.3226	0.6348	0.3383	0.1229	0.2348	0.7645
8.	Elkop	0.0621	0.0508	0.8109	0.0944	0.3678	0.2381	0.5030
9.	Energoaparatura	0.1729	0.0674	0.9778	0.0453	0.5348	0.8982	0.4477
10.	Energomontaż Południe	0.3317	0.0330	0.9672	0.0270	0.3736	0.8396	0.1969
11.	Energomontaż Północ	0.1473	0.2151	0.8241	0.1248	0.2556	0.4205	0.3489
12.	Energopol	0.2345	0.0851	0.9660	0.0537	0.3762	0.8351	0.1908
13.	Hydrobudowa Śląsk	0.1848	0.0099	0.9671	0.0208	0.5706	0.8429	0.3758
14.	Instal Kraków	0.4398	0.0378	0.9679	0.0141	0.5373	0.8462	0.1242
15.	Instal Lublin	0.2148	0.2416	0.8576	0.0832	0.3584	0.4260	0.3690
16.	Instal Export	0.2228	0.4565	0.9660	0.1962	0.1087	0.8466	0.3407
17.	Mitex	0.3457	0.1166	0.9733	0.0419	0.4329	0.8673	0.4133
18.	Mostostal Płock	1.0000	0.0593	0.9408	0.0051	0.3414	0.7596	0.0000
19.	Mostostal Warszawa	0.1909	0.0753	0.8979	0.0602	0.2770	0.6237	0.1812
20.	Mostostal Export	0.2034	0.4935	0.6724	0.3831	0.0000	0.6026	0.2997
21.	Naftobudowa	0.2153	0.1037	1.0000	0.0704	0.4851	1.0000	0.3798
22.	Pekabex	0.0434	0.1197	0.8179	0.1454	0.2824	0.3697	0.6153
23.	Pemug	0.0645	0.1627	0.9808	0.1683	0.3546	0.8903	0.8624
24.	Polnord	0.1509	0.1290	0.9660	0.0731	0.3033	0.8343	0.4409
25.	Prochem	0.3387	0.0762	0.9717	0.0326	0.3660	0.8568	0.0784
26.	Projprzem	0.4967	0.0000	0.9801	0.0000	0.5991	0.9182	0.0857
27.	Warbud	0.2345	0.0504	0.9701	0.0183	1.0000	0.8775	0.4199

Source: Own presentation

The data set prepared this way is ready to be used for calculating distances between particular objects. An abridged Euclidean distances matrix is presented in Table 5.

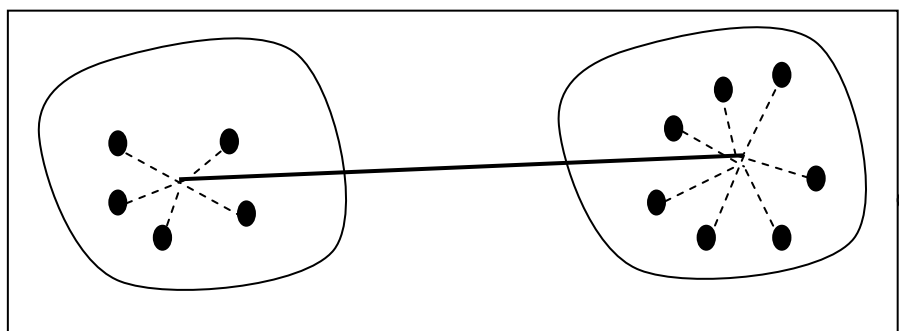
**Tab. 5: Euclidean distances matrix (fragment)**

<b>Company</b>	<b>Bick</b>	<b>Budimex</b>	<b>Budopol</b>	<b>(...)</b>	<b>Prochem</b>	<b>Projprzem</b>	<b>Warbud</b>
<b>Bick</b>		0.191	0.914	(...)	0.492	0.771	1.018
<b>Budimex</b>	0.191		0.772	(...)	0.300	0.580	0.859
<b>Budopol</b>	0.914	0.772		(...)	0.604	0.514	0.162
<b>(...)</b>	(...)	(...)	(...)	(...)	(...)	(...)	(...)
<b>Prochem</b>	0.492	0.300	0.604	(...)		0.282	0.643
<b>Projprzem</b>	0.771	0.580	0.514	(...)	0.282		0.479
<b>Warbud</b>	1.018	0.859	0.162	(...)	0.643	0.479	

Source: Own presentation

The k-means algorithm consists in assigning each examined object to a cluster having nearest centroids. A method is conducted in an iterative process. The aim is to partition a set  $P$  into k-groups so that companies within each cluster differ a little. A number of clusters or a number of iterations may be determined by a researcher. The latter can be performed, however, until value of criterion function does not increase. Criterion function expresses proportion between external group diversification (a sum of distances of groups' centroids from the centroid of all instances) to inner group diversification (a sum of distances of objects belonging to particular group from its center of mass).

Classical approach takes into account three steps. Initially, the set of objects should be divided into k-initial clusters. Then the centres of mass are computed for tentatively fixed groups of instances. The algorithm considers each object and assigns it to the cluster whose centroid is the closest. Afterwards, centres of mass for each cluster are recalculated after each next object is assigned. The process is iterated until no more assignment takes place [See: Panek, 2009, p.129).

**Fig. 2: Centroid method**

Source: Own presentation

## Results of analysis

Research results have been developed using *Statistica* software, including modification of k-means method consisting in exchange of particular pairs of instances within the various clusters. Classification was established after four iterations with assumed partition into five groups, following *W. Tarczyński's* recommendations that a final number of clusters should be equal to  $\sqrt{n}$ , where  $n$  is a number of instances (See: Łuniewska, Tarczyński, 2006, p. 56). The results are presented in the Table 6. Classification seems to be accurate, since five of nine distressed companies (*Bick*, *Elektromontaż Warszawa*, *Elektromontaż Export*, *Energomontaż Północ*, *Pekabex*) were assigned to the cluster comprising objects of 'the poorest' financial situation. The other distressed companies were classified as objects of 'weak' financial condition.

**Tab. 6: Results of the classification**

Classification	Companies	Group's size
The poorest financial situation	<i><u>Bick</u></i> , <i>Budimex</i> , <i><u>Elektromontaż Warszawa</u></i> , <i><u>Elektromontaż Export</u></i> , <i><u>Energomontaż Północ</u></i> , <i>Instal Export</i> , <i>Mostostal Export</i> , <i><u>Pekabex</u></i>	8
Weak financial situation	<i>Elektrobudowa</i> , <i><u>Elkop</u></i> , <i><u>Energoparatura</u></i> , <i>Energopol</i> , <i>Hydrobudowa Śląsk</i> , <i><u>Instal Lublin</u></i> , <i>Mostostal Warszawa</i> , <i><u>Naftobudowa</u></i> , <i>Pemug</i> , <i>Polnord</i>	10

Classification	Companies	Group's
Acceptable financial situation	<i>Mostostal Płock</i>	1
Good financial situation	<i>Budopol, Warbud</i>	2
The best financial situation	<i>Echo, Energomontaż Południe, Instal Kraków, Mitex, Prochem, Projprzem</i>	6

Source: Own presentation

It is worth mentioning that among twenty seven companies examined only one third were classified as those of ‘acceptable’, ‘good’ or ‘the best’ financial situation. Two thirds of the sample was assigned to the ‘weak’ and ‘the poorest’ clusters of which a half of objects had collapsed or had made agreements with creditors. It means that the rest – one third of the sample – emitted alarming signals of crisis even though they had not failed, in fact.

All these considerations have inclined towards looking for a response to the following question: what were the levels of financial ratios characterizing each cluster and its representatives? (See: Table 7).

**Tab. 7: Medium level of financial ratios for each cluster**

Classification	X <sub>1</sub>	X <sub>2</sub> [days]	X <sub>3</sub> [%]	X <sub>4</sub> [days]	X <sub>5</sub>	X <sub>6</sub> [%]	X <sub>7</sub>	Group's size
The best financial situation	2.03	110	1.92	126	1.28	2.35	0.41	6
Good financial situation	1.01	107	1.51	131	2.75	4.13	0.81	2
Acceptable financial situation	4.48	102	-5.63	58	1.14	-5.22	0.18	1
Weak financial situation	1.05	132	-5.94	177	1.32	-6.31	0.67	10
<b>* distressed companies</b>	<b>1.02</b>	<b>139</b>	<b>-12.45</b>	<b>181</b>	<b>1.39</b>	<b>-13.92</b>	<b>0.70</b>	<b>4</b>

Classification	X <sub>1</sub>	X <sub>2</sub> [days]	X <sub>3</sub> [%]	X <sub>4</sub> [days]	X <sub>5</sub>	X <sub>6</sub> [%]	X <sub>7</sub>	Group's size
The poorest financial situation	0.76	296	-57.56	581	0.62	-25.34	0.76	8
<b>* distressed companies</b>	<b>0.55</b>	<b>301</b>	<b>-78.52</b>	<b>668</b>	<b>0.71</b>	<b>-37.45</b>	<b>0.92</b>	<b>5</b>

Source: Own presentation

An average current ratio (X<sub>1</sub>) in the group of objects characterized by 'the poorest' financial situation reached a level of 0.76. However, the companies that had failed or made an agreement achieved a value equal to 0.55. That was a visible sign of deteriorating financial liquidity, since their current assets had covered short-term liabilities only in a half. The lowest volume of ratio was registered for *Bick*.

Medium level of return on sales (X<sub>3</sub>) in the group of 'the poorest' objects was equal to -57.6%, what had been caused by high losses on sales of three companies: *Bick* (-225.0%), *Elektromontaż Export* (-77.0%) oraz *Mostostal Export* (-68.2%). Despite the fact that firms assigned to this cluster had generated losses as a rule it is worth remarking that two of them – *Budimex* and *Instal Export* – had noted negligible profitability ratio. Moreover these two companies as the only in the cluster had not declared bankruptcy or made an agreement within the following two years.

The assessment of average collection period (X<sub>2</sub>) confirmed that efficiency of receivables' execution had been poor. The average ratio for the distressed companies reached approximately ten months. However *Bick*, *Mostostal Export* and *Instal Export* significantly exceeded this level. The average collection period amounted to more than two years in case of *Bick* and about a year in case of *Mostostal Export* and *Instal Export*. Similarly, unfavorable tendency was remarkable during observation of liabilities' payment time which amounted to over one and a half a year for 'the poorest' group of objects. To compare this period for 'good' and 'the best' companies was about four months.

The cluster including objects with the best financial situation comprised six companies: *Echo*, *Energomontaż Południe*, *Instal Kraków*, *Mitex*, *Echo*, *Prochem* i *Projprzem*. The current ratio incontestably decided about categorization, though, in the group of objects representing 'good' financial situation some ratios reached comparable levels. The features

that distinguished ‘the best’ companies from ‘weak’ and ‘the poorest’ entities concerned good financial liquidity (2.03), shorter average collection period (about three months) and liabilities payment time (about four months) as well as higher profitability of assets (2.35%) and sales (1.92%).

## **Conclusion**

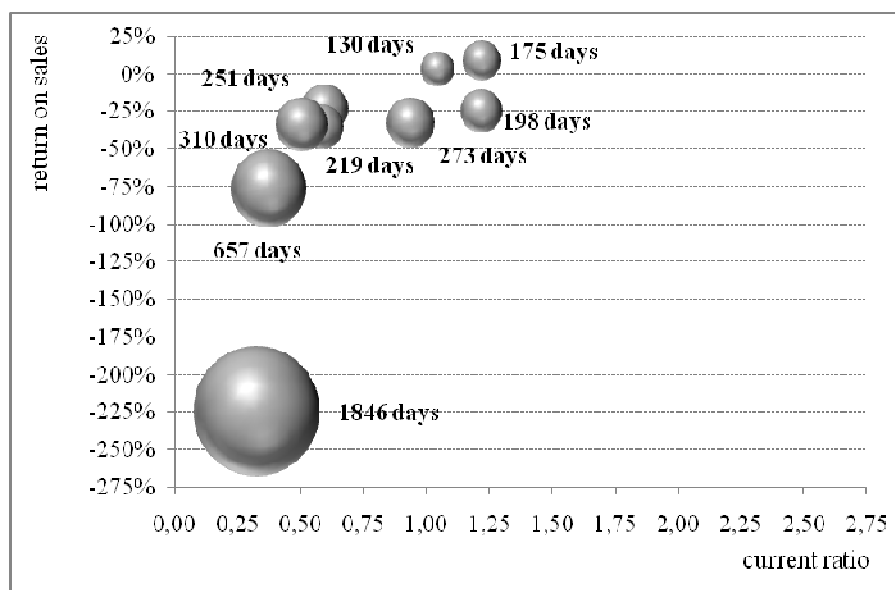
Application of k-means method provided promising results, since classification of objects was correct. It was also possible to determine levels of attributes, characterizing particular clusters regarding companies’ financial situation.

Although advantages of using k-means method should be appreciated, some existing limitations should not be forgotten. Deficiencies stem from time frame and researched object.

Firstly, it may be, sometimes, unreliable to compare results of classification made in various research periods. The main reasons that affect correct classification are changeable market conditions. Assuming that each year research sample is divided into five clusters representing financial condition of companies, it cannot be guaranteed that financial ratios building up clusters are always the same or – even if they were so – that they reflect the same levels.

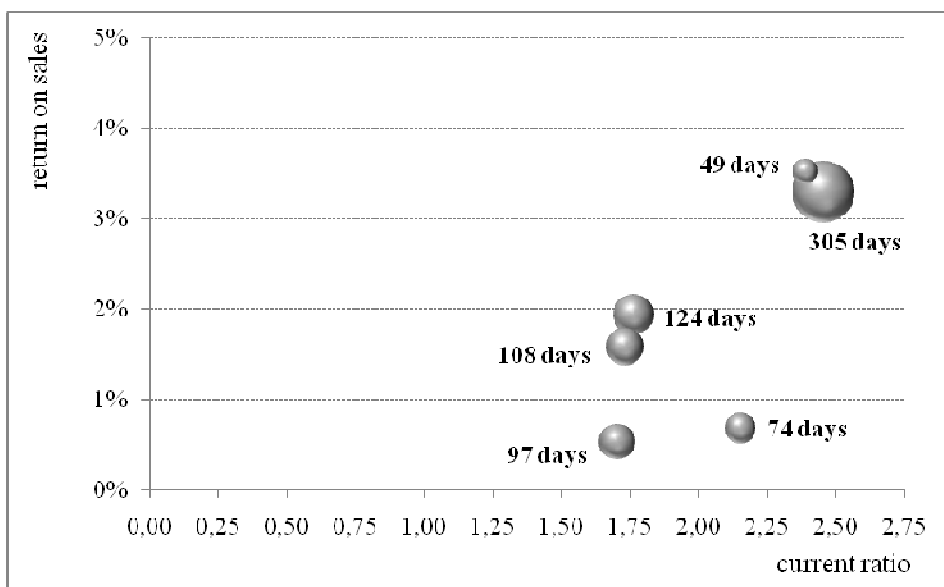
The next limitation concerns research objects – in this case construction companies publicly quoted at Warsaw Stock Exchange. Both ‘the best’ and ‘the poorest’ enterprises are chosen from the determined sample. It may lead to a false impression that companies with ‘the best’ financial condition represents desired volumes of ratios. In fact they are computed only for the narrow group of firms belonging to the construction sector, which was affected by the crisis. Therefore even ‘the best’ group of objects should not be unambiguously treated as those of excellent financial condition. In this particular case the examined companies which declared bankruptcy or made agreements with creditors in 2004-2005 had emitted clear signals of crisis at least a year before they collapsed. The signs had been visible especially in the financial liquidity sphere (See Figures 1-2). Current ratio did not exceed the level of 1.5 for each distressed company, whereas in case of ‘the best’ firms reached the level between 1.7 and 2.5. Most of the distressed enterprises were unprofitable and faced problems in setting their accounts what contrasted with ‘the best’ companies (except for *Echo*).

**Fig. 3: Current ratio, return on sales and liabilities payment time for distressed companies**



Source: Own presentation

**Fig. 4: Current ratio, return on sales and liabilities payment time for the best companies**



Source: Own presentation

The aforementioned observations should make investors, which are interested in achieving high return on their capital investments, sensible that monitoring financial condition of particular company must not be done in isolation from the situation in the branch.

## References

- [1] Aldenderfer, M. S. – Blashfield, R. K. (1984): *Cluster Analysis*. Sage University, 1984.
- [2] Becla, A. – Zielińska, A. (2003): *Elementy statystyki i metod ilościowych*, Wrocław, I-BIS Usługi Komputerowe, 2003.
- [3] Dunn, G. – Everitt, B. S. (1982): *An Introduction to Mathematical Taxonomy*. Cambridge University, 1982.
- [4] Dyczkowska, J. (2007): *O wykorzystaniu metody sum standaryzowanych w diagnozowanie kryzysu przedsiębiorstwa*. Zeszyty Teoretyczne Rachunkowości, 2007, Tom 41 (97), p. 67-79.
- [5] Dziechciarz, J. (ed.) (2003): *Ekonometria. Metody, przykłady, zadania*. Wrocław, Wydawnictwo Akademii Ekonomicznej, 2003.
- [6] Kolenda, M. (2006): *Taksonomia numeryczna. Klasyfikacja, porządkowanie i analiza obiektów wielocechowych*. Wrocław, Prace Naukowe Akademii Ekonomicznej we Wrocławiu Nr. 1106, 2006.
- [7] Kukuła K. (2000): *Metoda unitaryzacji zerowanej*. Warszawa, PWN, 2000.
- [8] Łuniewska M., Tarczyński W. (2006): *Metody wielowymiarowej analizy porównawczej na rynku kapitałowym*. Warszawa, PWN, 2006.
- [9] Nowak E. (1990): *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*. Warszawa, PWE, 1990.
- [10] Nowak E. (1997): *Zarys metod ekonometrii*. Warszawa, PWN, 1997.
- [11] Ostasiewicz W. (1999): *Statystyczne metody analizy danych*. Wrocław, Wydawnictwo Akademii Ekonomicznej, 1999.
- [12] Panek T. (2009): *Statystyczne metody wielowymiarowej analizy porównawczej*, Warszawa, Oficyna Wydawnicza SGH, 2009.
- [13] Pawłowicz L. (ed.) (2001): *Ekonomika przedsiębiorstw*. Gdańsk, Ośrodek Doradztwa i Doskonalenia Kadr, 2001.



- [14] Romesburg H. C. (2004): *Clustering Analysis for Researchers*. North Carolina, Lulu Press, 2004.
- [15] Silva L. A. L. – Campbell, J. A. – Eastaugh N. – Buxton B. F. (2008): *A Case for Numerical Taxonomy in Case-Based Reasoning*. In: *Advances in Artificial Intelligence – SBIA 2008*. Berlin, Heidelberg, Springer Verlag, 2008.
- [16] Sojak, S. – Stawicki, J. (2001): *Wykorzystanie metod taksonomicznych do oceny kondycji ekonomicznej przedsiębiorstw*. Zeszyty Teoretyczne Rachunkowości, 2001, Tom 3 (59), p. 56-67.
- [17] Witkowska, D. (2002): *Sztuczne sieci neuronowe i metody statystyczne*. Warszawa, C. H. Beck, 2002.

## **Usefulness of K-means Method in Detection Corporate Crisis**

*Joanna DYCZKOWSKA*

### **ABSTRACT**

Market situation and business environment of construction companies influence significantly decisions met by this group of entities. These decisions are reflected in financial statements, later on. The evaluation of financial condition, which aims at diagnosing corporate crisis, must not disregard a market situation. Taking this assumption into account a classification of publicly quoted construction companies using k-means method was conducted. This procedure enabled to divide the examined sample into five clusters of companies characterized by 'the best', 'good', 'acceptable', 'weak' and 'the poorest' financial condition. The application of the aforementioned algorithm helped also to determine levels of financial ratios typical for each cluster. This kind of analytical approach is particularly useful for investors, since it informs how particular companies perform in comparison to their competitors.

**Key words:** Numerical taxonomy; Cluster analysis; K-means method; Financial situation; Current ratio; Average collection period; Return on sales ratio; Liabilities payment time; Assets turnover; Return on assets; Debt ratio; Coefficient of variation; Pearson correlation coefficient; Euclidean distance.

**JEL classification:** G32, G33, M41.