

## Cluster analysis in retail segmentation for credit scoring

Sanja Scitovski<sup>1,\*</sup> and Nataša Šarlija<sup>2</sup>

<sup>1</sup> *Josip Juraj Strossmayer University of Osijek*  
*Trg Sv. Trojstva 3, 31 000 Osijek, Croatia*  
*E-mail: {sscitov@unios.hr}*

<sup>2</sup> *Faculty of Economics, Josip Juraj Strossmayer University of Osijek*  
*Trg Ljudevita Gaja 7, 31 000 Osijek, Croatia*  
*E-mail: {natasa@efos.hr}*

**Abstract.** The aim of this paper is to segment retail clients by using adaptive Mahalanobis clustering in a way that each segment can be suitable for separate credit scoring development such that a better risk assessment of retail clients could be accomplished. A real data set on retail clients from a Croatian bank was used in the paper. Grouping of the data point set is carried out by using the adaptive Mahalanobis partitioning algorithm (see, e.g., [20]). It is an incremental algorithm, which recognizes ellipsoidal clusters with the main axes in the directions of eigenvectors of the corresponding covariance matrix of the data set. On the basis of the given data set, by using the well-known DIRECT algorithm for global optimization it is possible to search successively for an optimal partition with  $k = 2, 3, \dots$  clusters. After that, a partition with the most appropriate number of clusters is determined by using various validity indexes. Based on the description of each cluster, banks could decide to develop a separate credit scoring model for each cluster as well as to create a business strategy customized to each cluster.

**Key words:** cluster analysis; credit scoring; segmentation; data mining; adaptive Mahalanobis clustering; most appropriate number of clusters; data classification

Received: September 22, 2014; accepted: December 2, 2014; available online: December 30, 2014

---

## 1. Introduction

Cluster analysis is a method frequently used in client segmentation. It is the process of dividing clients into well separated homogenous groups. The idea is to create customized marketing strategies for selected segments in order to satisfy clients' needs better. Various methods have been used for market segmentation, such as different clustering methods, fuzzy methods, regression methods, neural networks, self-organizing maps, and others (see, e.g., [4, 8, 19, 22]). Classical cluster analysis methods give an excellent level of precision [14]. Cluster analysis can also be used for credit scoring. It is a process of determining how likely applicants are to default with their repayments [10]. The aim is to assess the risk of default associated with a credit product/decision.

---

\*Corresponding author.

However, it is not very often that cluster analysis is used for credit scoring. Usually, one credit scoring model is built for the entire client population. Sometimes the population can be heterogeneous so it is reasonable to make segmentation of the population and then for each segment develop separate scoring models. [1] emphasized various factors that can trigger a decision to build more than one scoring model: marketing, customer, data, process and model fit factor. Since building of multiple scorecards requires additional costs due to development, implementation, maintenance, monitoring and validation, these should be compensated by improvement in performance. [13] stated that segmentation usually increases performance by 5 to 10 percent. But, it is not always the case. In their research, [3] did not show any improvement in performance.

According to [3], there are several approaches to using segmentation in credit scoring: (i) Making clusters by hierarchical clustering followed by making scoring models; (ii) Logistic regression follows Classification and Regression Trees (CART) or Chi-squared Automatic Interaction Detection (CHAID) trees; (iii) Multi-level segmentation approach: combining experience-based segmentation (at higher levels) and segmentation based on a dedicated score (at lower levels), (iv) simultaneous methods proposed by [9]; (v) Fair Isaacs Adaptive Random Trees (ART) technology. In contrast to these approaches, in our research grouping of the data point set in ellipsoidal clusters by using adaptive Mahalanobis clustering describes the nature of the problem better, since clusters in the optimal partition very rarely take a spherical form.

## 2. Cluster analysis in credit scoring

### 2.1. Previous research

Previous research shows that there are two streams of papers that use cluster analysis in credit scoring. In one of them, cluster analysis is used as one of the classification methods for classifying good and bad applicants. In the other stream, cluster analysis is used as a method for segmenting clients into clusters on which separate scoring models are built.

[6] compared results of several classification methods to cluster creditworthy customers against non credit worthy ones. They compared Time Adaptive Self-Organizing Map Neural Network (TASOM), K-means, Partitioning around Medoids (PAM) and Self-Organizing Map (SOM) and realized that TASOM has the best performance in clusters of customers.

[3] made segmentation of clients' data sets using CART and CHAID and then built scorecards for the identified segments. They also applied a simultaneous approach where the LOTUS (Logistic Trees with Unbiased Selection) algorithm was used to develop both segmentation and scorecards. They are both optimized at the same time. Additionally, a single-scorecard model was estimated so it can be compared to multi-scorecard models. All scorecards were built by using logistic regression with stepwise selection. Results showed that for none of the analyzed real-world data sets, multi-scorecard models perform considerably better than logistic regression.

In [24], the combination of cluster analysis and decision tree models is investigated. In the first stage, bank customers are segmented into clusters that are characterized by similar features and then, in the second step, for each group, decision trees are built to obtain rules that may indicate clients expected not to repay the loan. Results obtained on the real credit risk data sets showed higher precision and simplicity of rules obtained for each cluster than for rules connected with the whole data set.

[12] proposed a behavioral scoring model, which dynamically accommodates the changes of borrowers' characteristics after the loans are made. To increase the prediction efficiency, they segmented the data set into three clusters by using the K-means algorithm. On each cluster, a scoring classifier is built. They developed a cluster-based dynamic scoring and a non-cluster scoring model. The results showed that clustering lowered the misclassification rate of the single rule model. Additionally, each of three clusters generally showed a lower misclassification rate than the single rule model.

## 2.2. Data clustering

Clustering or grouping a data set  $\mathcal{A} = \{a^i \in \mathbb{R}^n: i = 1, \dots, m\}$  with  $n$  features in several compact and well separated clusters has practical importance in a wide variety of applications such as biology, medicine, physics, environmental science, business, the facility location problem (see, e.g., [2, 20]). The problem is as follows: the set  $\mathcal{A}$  should be partitioned into  $1 \leq k \leq m$  nonempty disjoint subsets  $\pi_1, \dots, \pi_k$ , such that

$$\bigcup_{i=1}^k \pi_i = \mathcal{A}, \quad \pi_r \cap \pi_s = \emptyset, \quad r \neq s, \quad |\pi_j| \geq 1, \quad j = 1, \dots, k. \quad (1)$$

Subsets  $\pi_1, \dots, \pi_k$  are called *clusters in*  $\mathbb{R}^n$ ; the set of all clusters is called a partition and it will be denoted by  $\Pi = \{\pi_1, \dots, \pi_k\}$ . The collection of all such partitions will be denoted by  $\mathcal{P}(\mathcal{A}, k)$ . Suppose also that a weight  $w_i > 0$  is associated to each data point  $a^i \in \mathcal{A}$ .

First, if components of data points are not of equal range, they should be normalized [7, 20]. Since  $\mathcal{A}$  is a finite set, one can always suppose that there exists a hyperrectangle  $[\alpha, \beta] = \{x \in \mathbb{R}^n: \alpha_s \leq x_s \leq \beta_s\}$ , where  $\alpha = (\alpha_1, \dots, \alpha_n)^T$ ,  $\beta = (\beta_1, \dots, \beta_n)^T \in \mathbb{R}^n$ , such that  $\mathcal{A} \subset [\alpha, \beta]$ . By using the mapping  $T: [\alpha, \beta] \rightarrow [0, 1]^n$ , where

$$T(x) = D(x - \alpha), \quad D = \text{diag} \left( \frac{1}{\beta_1 - \alpha_1}, \dots, \frac{1}{\beta_n - \alpha_n} \right), \quad (2)$$

the hyperrectangle  $[\alpha, \beta]$  is transformed into the hypersquare  $[0, 1]^n$  and the set  $\mathcal{A}$  into set  $\mathcal{B} = \{T(a^i): a^i \in \mathcal{A}\} \subset [0, 1]^n$ .

After clustering the set  $\mathcal{B}$ , the obtained results will be transformed back into the hyperrectangle  $[\alpha, \beta]$  by the mapping  $T^{-1}: [0, 1]^n \rightarrow [\alpha, \beta]$ , where

$$T^{-1}(x) = D^{-1}x + \alpha. \quad (3)$$

If we introduce some distance-like function (see, e.g., [11])  $d: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+$ ,  $\mathbb{R}_+ := [0, +\infty)$ , then to each cluster  $\pi_j \in \Pi$  its center  $c_j$  defined by

$$c_j = \operatorname{argmin}_{x \in \mathbb{R}^n} \sum_{a \in \pi_j} d(x, a) \quad (4)$$

can be associated. Now, a globally optimal  $k$ -partition  $\Pi^* \in \mathcal{P}(\mathcal{A}, k)$  can be defined as

$$\Pi^* = \operatorname{argmin}_{\Pi \in \mathcal{P}(\mathcal{A}, k)} \mathcal{F}(\Pi), \quad \mathcal{F}(\Pi) = \sum_{j=1}^k \sum_{a \in \pi_j} d(c_j, a), \quad (5)$$

where  $\mathcal{F}: \mathcal{P}(\mathcal{A}, k) \rightarrow \mathbb{R}_+$  is the objective function.

Since very elongated elliptical clusters are expected in our research, the adaptive Mahalanobis distance-like function

$$d_M^j(u, v; S) = \sqrt[3]{\det S_j} (u - v)^T S_j^{-1} (u - v) \quad (6)$$

induced by a set of data points from the cluster  $\pi_j$  will be used [2, 5], where

$$S_j = \frac{1}{W_j} \sum_{a^i \in \pi_j} w_i (c_j - a^i)(c_j - a^i)^T, \quad W_j = \sum_{a^i \in \pi_j} w_i, \quad (7)$$

is a *covariance matrix* and  $c_j$  is the centroid of the cluster  $\pi_j$ . It can be shown that [2, 18]

$$c_j = \frac{1}{W_j} \sum_{a^i \in \pi_j} w_i a^i, \quad W_j = \sum_{a^i \in \pi_j} w_i. \quad (8)$$

From a geometric point of view, in this case the partition  $\Pi$  consists of elliptical shape clusters that are as compact and relatively strongly separated as possible.

Searching for an optimal partition is a very demanding numerical procedure. In this paper, we adopt a new fast incremental partitioning algorithm proposed in [15, 17], which is able to find either a globally optimal partition or a locally optimal partition of the set  $\mathcal{A} \subset \mathbb{R}^n$  close to the global one. This algorithm uses the well-known DIRECT algorithm for global optimization [7] and successively gives optimal partitions for each  $k = 2, \dots, k_{max}$ , where  $k_{max}$  is the maximum number of clusters that makes sense to be calculated. Therefore, it is very suitable for searching for a partition with the most appropriate number of clusters.

Since in this research the number of clusters is not given in advance, it is natural to search for an optimal partition which consists of clusters that are as compact and relatively strongly separated as possible. This can be done by using some of well-known indexes [6, 21]: Calinski-Harabasz (CH), Davies - Bouldin (DB), Simplified Silhouette Width Criterion (SWC), Hypervolume (HV) index. More compact and better separated clusters in an optimal partition will result in a greater CH index, a smaller DB index, a greater SWC index and a smaller HV index. By applying the aforementioned indexes a partition with the most appropriate number of clusters will be determined.

### 3. Data and variables

The data set for this research consists of 3165 retail clients of one bank in Croatia. These are consumer loans for the purchase of goods. Available variables for this research are grouped into 5 groups with appropriate features:

- (i) Demographic: client age (Mean= 39.1, SD= 12.6);
- (ii) Behavioral: payment method (Paying slip 64.2%; Standing order 35.8%); purchased goods (White goods 17%; Brown goods 29.8%; Home supplies 5.4%; Communication devices 22.1%; Furniture 7.6%; Computers 3.3%; Other 14.8%); relationship with the client (New client 89.1%; Old client 10.9%);
- (iii) Personal: occupation (Entrepreneur 12.7%; Pensioner 11.9%; Employee 75.4%), length of service in months (Mean= 71.77, SD = 78.73), marital status (Single 21.2%; Married 60.4%; Widow(er) 13.6%; Divorced 4.8%), housing (Rented 9.5%; Tenancy right 4.3%; Apartment 42%; House 15.6%; With parents 20.2%; Other 8.4%);
- (iv) Financial: wage (monthly salary of a client in HRK) (Mean= 3735.6 SD = 2587.6), payment in cash (Mean= 837.8, SD = 1283.2);
- (v) Loan: good price (Mean= 4740.9 SD = 3184.6), installment (Mean= 406.2, SD = 311.1), status (Good 86.7%; Bad 13.3%).

**Remark 1.** Note that e.g. the feature “purchased goods” can take only 7 different values:  $1, \dots, 7$ . Therefore, during the clustering process the appearance of a cluster with all elements having the same value (e.g., “1”) for the feature “purchased goods” can be expected.

The following lemma shows what consequences that can have on the clustering process.

**Lemma 1.** Let  $\mathcal{A} = \{a^i = (a_1^i, \dots, a_n^i) \in \mathbb{R}^n : i = 1, \dots, m\}$  be the set of data points in  $\mathbb{R}^n$  with corresponding weights  $w_i > 0$  and let  $c = (c_1, \dots, c_n) \in \mathbb{R}^n$  be the centroid with the components

$$c_s = \frac{1}{W} \sum_{i=1}^m w_i a_s^i, \quad s = 1, \dots, n, \quad W = \sum_{i=1}^m w_i.$$

(I) The covariance matrix

$$S = \frac{1}{W} \sum_{i=1}^m w_i (a^i - c)(a^i - c)^T = \frac{1}{W} \begin{bmatrix} \sum w_i (a_1^i - c_1)^2 & \cdots & \sum w_i (a_1^i - c_1)(a_n^i - c_n) \\ \vdots & \ddots & \vdots \\ \sum w_i (a_n^i - c_n)(a_1^i - c_1) & \cdots & \sum w_i (a_n^i - c_n)^2 \end{bmatrix},$$

is positive definite if and only if the set of vectors

$$\varphi_s = (a_s^1 - c_s, \dots, a_s^m - c_s) \in \mathbb{R}^m, \quad s = 1, \dots, n, \quad (9)$$

is linearly independent.

(II) The covariance matrix  $S$  is singular

- (i) if  $m < n$ , or
- (ii) if there exists  $s_0 \in \{1, \dots, n\}$ , such that  $a_{s_0}^1 = \dots = a_{s_0}^m$ , or
- (iii) if there exist  $r_0, s_0 \in \{1, \dots, n\}$ , such that  $a_{r_0}^i = a_{s_0}^i, \forall i = 1, \dots, m$ .

**Proof.** In order to prove claim (I), note that the covariance matrix  $S$  is the Gram matrix of the vectors  $\varphi_1, \dots, \varphi_n$  with the usual weighted scalar product. The proof of claim (I) follows from the fact that the Gram matrix is positive definite if and only if the vectors  $\varphi_1, \dots, \varphi_n$  are linearly independent.

If  $m < n$ , (9) is the set of  $n$  vectors from the vector space of dimension less than  $n$ , and therefore, it is linearly dependent.

If all data points  $a^i \in \mathcal{A}$  have the  $s_0$ -th component equal to  $\kappa \in \mathbb{R}$ , then  $c_{s_0} = \kappa$ , and the vector  $\varphi_{s_0}$  is the zero-vector. Therefore, set (9) is linearly dependent.

If  $r_0$ -th and  $s_0$ -th components of all data points  $a^i \in \mathcal{A}$  are equal, then  $c_{r_0} = c_{s_0}$ . Therefore,  $\varphi_{r_0} = \varphi_{s_0}$ , and set (9) is linearly dependent.

Since in all mentioned cases, set (9) is linearly dependent, the corresponding Gram matrix is singular. This means that the covariance matrix  $S$  is singular, too.  $\square$

By using Lemma 1, the problems mentioned in Remark 1 can practically be solved such that the small perturbation of the component  $s_0$  of all data points  $a^i \in \mathcal{A}$  is done

$$a_{s_0}^i := a_{s_0}^i + \epsilon_i, \quad \epsilon_i \sim \mathcal{U}(\alpha, \beta),$$

where  $\alpha < \beta$  are significantly less than  $|a_{s_0}^i|$ . For example, instead of number “1”, by the feature “purchased goods” it is sufficient to put uniform distributed numbers from the interval  $[1.0001, 1.001]$ . Then the vector  $\varphi_{s_0}$  will not be the zero vector any more, the covariance matrix will not be singular any more, and the clustering process will be carried out further without any problems.

## 4. Results

Different clustering of data with different numbers of features was carried out during research. First, we give the clustering example based on two features (Age and Loan) because in this case, in addition to the numerical characteristic, we can consider geometry of the problem, too. We also give the clustering example based on five features (Age, Length of Service, Payment in Cash, Salary, Purchased Goods). In this case, we will mention a statistical analysis of obtained clusters.

### 4.1. Clustering based on two features: Age and Loan

Let us consider the data set  $\mathcal{A} = \{a^i = (x_i, y_i) \in \mathbb{R}^2 : i = 1, \dots, 3165\}$  with two features:  $x_i$  (Age) and  $y_i$  (Loan). By using transformation (2) the normalized set  $\mathcal{B}$  is obtained. The algorithm proposed in [15] is applied to the data set  $\mathcal{B}$ , and optimal partitions with  $2 \leq k \leq 6$  clusters are determined. Figure 1 shows optimal partitions with  $k=3, 4$  and 5 clusters.

$k$	SWC	DB	CH ( $\times 10^2$ )	HV ( $\times 10^{-8}$ )
2	<b>0.78</b>	0.52	37.2	16.2
3	0.76	0.50	<b>53</b>	<b>8.5</b>
4	0.75	0.48	50	9.3
5	0.73	<b>0.47</b>	37	117

Table 1: *Indexes of optimal partitions. The best performers for each index are bolded.*

Most of the indexes indicate that the optimal partition with three clusters is the most appropriate partition (see Table 1), although the DB index indicates that it could be an optimal partition with five clusters, too.

By using the inverse transformation (3) the results for the original data set  $\mathcal{A}$  are obtained.

Furthermore, we consider the optimal partition with three clusters  $\Pi(\pi_1, \pi_2, \pi_3)$ . Each cluster  $\pi_j$  of this partition is characterized by its centroid  $c_j$ , and the corresponding covariance matrix  $S_j$ . To each cluster we associate a Mahalanobis circle (M-circle)

$$K_j(c_j, \sigma_j) = \{a \in \mathbb{R}^n : d_M(c_j, a; cov_j) \leq \sigma_j^2\}, \quad \sigma_j^2 = \frac{1}{|\pi_j|} \sum_{a \in \pi_j} d_M(c_j, a; cov_j), \quad (10)$$

with center  $c_j$  (see ellipses in Figure 1a) and the radius equal to the Mahalanobis standard deviation of all data points from the cluster  $\pi_j$ . This circle is called the main M-circle.

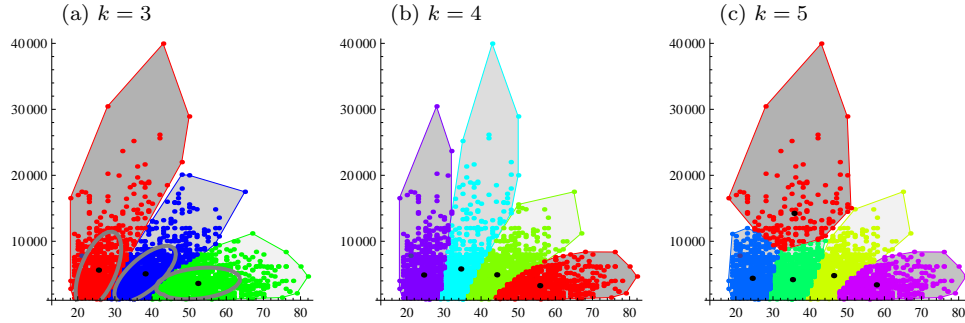


Figure 1: *Optimal partitions with  $k = 3, 4$  and 5 clusters. The corresponding main M-circle is associated to each cluster of a 3-partition.*

The main M-circle  $K_j$  consists of approximately 68% data points from the cluster  $\pi_j$ , and it is situated in the rectangle  $[x_L^{(j)}, x_R^{(j)}] \times [y_L^{(j)}, y_R^{(j)}]$ , where

$$x_L^{(j)} = \min\{x_i : (x_i, y_i) \in K_j \cap \pi_j\}, \quad x_R^{(j)} = \max\{x_i : (x_i, y_i) \in K_j \cap \pi_j\}, \quad (11)$$

$$y_L^{(j)} = \min\{y_i : (x_i, y_i) \in K_j \cap \pi_j\}, \quad y_R^{(j)} = \max\{y_i : (x_i, y_i) \in K_j \cap \pi_j\}. \quad (12)$$

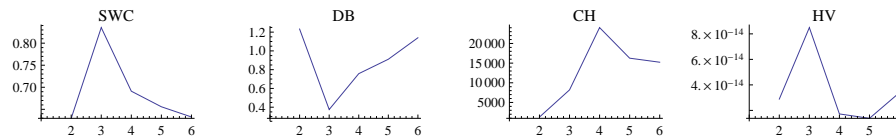
Table 2 shows the main characteristics of all clusters of the optimal partition  $\Pi$ : the number of data points  $|\pi_j|$ , the number of data points in the corresponding main M-circle  $|K_j|$ , the centroid  $c_j$ , Mahalanobis standard deviation  $\sigma_j$ , borders of features (11)–(12) of data points from the main M-circle  $K_j$  and eigenvalues of the corresponding covariance matrix  $\lambda_1^j > \lambda_2^j$ .

	$ \pi_j $	$ K_j $	$c_j$	$\sigma_j$	Borders of features		$\lambda_1^j > \lambda_2^j$
					Age	Loan	
$\pi_1$	1098	747	(52.3, 3612)	162.5	[41, 63]	[1380, 5996]	0.0160 > 0.0018
$\pi_2$	1070	771	(38.1, 5068)	158.5	[30, 46]	[1198, 9131]	0.0123 > 0.0021
$\pi_3$	997	747	(25.6, 5632)	172.3	[20, 31]	[1162, 11000]	0.0132 > 0.0027

Table 2: *The main characteristics of all clusters of the optimal partition  $\Pi$* 

## 4.2. Clustering based on five features: Age, Length of Service, Payment in Cash, Salary, Purchased Goods

Let us consider the data set  $\mathcal{A} = \{a^i \in \mathbb{R}^n : i = 1, \dots, 3165\}$  with  $n = 5$  features: Age, Length of Service, Payment in Cash, Salary, Purchased Goods. First, the normalized set  $\mathcal{B}$  is obtained by using transformation (2). The algorithm proposed in [15] is applied to the data set  $\mathcal{B}$  and optimal partitions with  $2 \leq k \leq 6$  clusters are obtained. As can be seen in Figure 2, the SWC and the DB index show that an optimal partition with 3 clusters is the most appropriate partition, although the CH index (resp. the HV index) indicates that it could be the optimal partition with four (resp. five) clusters, too.

Figure 2: *Indexes of optimal partitions for the data with five features*

In what follows, we consider the optimal 3-partition of the retail client  $\Pi(\pi_1, \pi_2, \pi_3)$  with 1642, 936, 587 data points, which are situated in the main M-circles as follows 1179, 630, 435.

The results are presented in Table 3. The first cluster is characterized by clients who are on average 39 years old and have 6 years of working experience. Most of them are employees. They are married and have their own apartment. Compared to other clusters, this one has the lowest proportion of clients with tenancy rights. They buy white and brown goods. The price of goods they buy is on average 4,800 HRK. They are prepared to pay 740 HRK in cash. The proportion of bad clients in the first cluster is 13.15%.

The second cluster contains the youngest clients who are on average 36 years old and with 5 years of working experience. 75% of them buy communication related goods and the rest of them buy furniture. In comparison to other clusters, this one contains the highest proportion of singles with tenancy rights. They are willing to give the lowest amount of cash in advance, compared to clients in other clusters. They also buy the cheapest goods with the lowest installment. The proportion of bad clients in the second cluster is 13.14%.

The oldest clients are segmented in the third cluster. Most of them are employees but compared to other clusters, this one has the highest proportion of entrepreneurs. Also, most of them possess their own apartments. They buy other goods such as construction or renovation work material. The goods they buy are most expensive



and, compared to other clients, they are willing to give the highest amount of cash. In order to compare clusters, ANOVA test was performed for numerical variables and **chi-square test** for independence of categorical variables. It has been shown that the clusters are statistically significant according to clients' age ( $p < 0.0001$ ), purchased goods ( $p < 0.0001$ ), relationship between a client and the bank ( $p = 0.0463$ ), occupation ( $p < 0.0001$ ), marital status ( $p < 0.0001$ ), housing ( $p < 0.0001$ ), price of goods ( $p < 0.0001$ ) and installment ( $p < 0.0001$ ), but not according to the payment method ( $p = 0.1073$ ), salary ( $p = 0.6868$ ) and proportion of bad clients ( $p = 0.9161$ ).

Cluster	Age	Marital status (in %)					Housing (in %)				
		Sin.	Marr.	Wid.	Div.	Ren.	Ten.	Ap.	Hou.	With.	Other
$\pi_1$	39.7	20	60	15	5	10	4	41	15	21	9
$\pi_2$	36.5	27	57	13	3	10	5	37	21	19	8
$\pi_3$	41.5	16	67	11	6	6	5	52	9	21	7
Cluster	Occupation (in %)			Service			Purchased goods (in %)				
	Ent.	Pen.	Emp.	Months	White	Brown	Home	CD	Furn.	Comp.	Other
$\pi_1$	12	13	75	65.3	33	57	10	0	0	0	0
$\pi_2$	11	8	81	60.7	0	0.2	0.3	74.9	24.5	0.1	0
$\pi_3$	18	14	68	68.6	0	0	0	0	2	18	80
Cluster	Salary HRK	Cash HRK	Payment method (in %)		Previous experience (in %)		Client status (in %)				
			Wage	Cash	New	Old	Good	Bad			
$\pi_1$	3711	740	65	35	90	10	87	13			
$\pi_2$	3726	486	65	35	87	13	87	13			
$\pi_3$	3818	1672	60	40	91	9	86	14			

Table 3: Features of retail clients in clusters

## 5. Conclusions

Segmentation of retail clients by using adaptive Mahalanobis clustering was conducted in the paper. By this method, ellipsoidal clusters defined based upon the covariance matrix of the data set are determined from each cluster separately. In this way, the main M-circle is determined, too. Other methods mentioned in the paper mainly use the LS distance-like function, which produces spherical clusters. Therefore, the real nature of the problem is not taken into account sufficiently. In case when essentially ellipsoidal clusters are expected, adaptive Mahalanobis clustering gives results that are superior to other clustering methods. In addition, the method gives a satisfactory approximation of the globally optimal partition with  $k = 2, 3, \dots$  ellipsoidal clusters, and hence it is possible to detect a partition with the most appropriate number of clusters by using standard indexes.

The results showed that optimal partitioning of the data set  $\mathcal{A}$  with 5 features has three clusters. The resulting clusters are significantly different to each other according to 8 variables. Each cluster has its own specificity and as such can be a separate segment for a unique credit risk strategy. It is obvious that a specific credit program can be designed for young clients who are usually single, buy communication related devices and are willing to give the smallest amount of cash in advance (second cluster) as opposed to older clients who buy most expensive goods for construction and renovation and are willing to give the highest amount of cash in advance (third cluster). As some previous research has already shown, it is reasonable to develop a separate scoring model for each segment. Performance can be increased by 5 to 10 percent. In order to test the performance, guidelines for further research are to

develop separate scoring models for each of the resulting clusters as well as for the whole data set and then to compare the performance of all four clusters.

## 6. Acknowledgement

The authors would like to thank anonymous referees and journal editors for their careful reading of the paper and insightful comments that helped us improve the paper.

## References

- [1] Anderson, R. (2007). *The Credit Scoring Toolkit*, Oxford University Press, New York
- [2] Bezdek, J.C., Keller, J., Krisnapuram, R. and Pal, N.R. (2005). *Fuzzy models and algorithms for pattern recognition and image processing*, Springer.
- [3] Bijak, K. and Thomas, L.C. (2012). Does segmentation always improve model performance in credit scoring?. *Expert Systems with Applications*, 39, 2433–2442.
- [4] Chan, K. and Kwong, C. (2012). Market segmentation and ideal point identification for new product design using fuzzy data compression and fuzzy clustering methods, 12, 1371–1378.
- [5] Durak, B. (2011). *A Classification Algorithm Using Mahalanobis Distances Clustering of Data with Applications on Biomedical Data Set*. Ph.D. thesis. The Graduate School of Natural and Applied Sciences of Middle East Technical University.
- [6] Gholamian, M. and Jahanpour, S. and Sadatrasoul, S. (2013). A new method for clustering in credit scoring problems. *Journal of Mathematics and Computer Science*, 6, 97–106.
- [7] Grbić, R., Nyarko, E.K. and Scitovski, R. (2013). A modification of the DIRECT method for Lipschitz global optimization for a symmetric function, *Journal of Global Optimization*, 57, 1193–1212.
- [8] Hanafizadeh, P. and Mirzazadeh, M. (2011). Visualizing market segmentation using self-organizing maps and fuzzy delphi method adsl market of a telecommunication company. *Expert Systems with Applications*, 38, 198–205.
- [9] Hand, D., Sohn, S. and Kim, Y. (2005). Optimal bipartite scorecards. *Expert Systems with Applications*, 29, 684–690.
- [10] Hand, D.J. and Henley, W.E. (1997). Statistical classification methods in consumer credit scoring: A review. *Journal Royal Statistical Society A*, 160, 523–541.
- [11] Kogan, J. (2007). *Introduction to Clustering Large and High-dimensional Data*. Cambridge University Press.
- [12] Lim, M.K. and Sohn, S.Y. (2007). Cluster-based dynamic scoring model. *Expert Systems with Applications*, 32, 427–431.
- [13] Makuch, W. (2001). The basics of a better application score, In: *Handbook of Credit Scoring*. Glenlake Publishing Company, Chicago, 127–148.
- [14] Mingoti, S.A. and Lima, J.O. (2006). Comparing som neural network with fuzzy c-means, k-means and traditional hierarchical clustering algorithms. *European Journal of Operational Research*, 174, 1742–1759
- [15] Morales-Esteban, A., Martínez-Álvarez, F., Scitovski, S. and Scitovski, R. (2014). A fast partitioning algorithm using adaptive mahalanobis clustering with application to seismic zoning. *Computers & Geosciences*, 73, 132–141.
- [16] Nievergelt, Y. (1994). Total least squares: state-of-the-art regression in numerical analysis. *SIAM Review*, 36, 258–264.

- [17] Scitovski, R. and Scitovski, S. (2013). A fast partitioning algorithm and its application to earthquake investigation. *Computers & Geosciences*, 59, 124–131.
- [18] Späth, H. (1983). *Cluster-Formation und Analyse*. R. Oldenburg Verlag, München
- [19] Šarlija, N., Benšić, M. and Zekić-Sušac, M. (2009). Comparison procedure of predicting the time to default in behavioural scoring. *Expert Systems with Applications*, 36, 8778–8788.
- [20] Theodoridis, S. and Koutroubas, K. (2009). *Pattern Recognition*. Academic Press, Burlington, 4<sup>th</sup> edition.
- [21] Vendramin, L., Campello, R.J.G.B. and Hruschka, E.R. (2009). On the comparison of relative clustering validity criteria, In: *Proceedings of the SIAM International Conference on Data Mining, SDM 2009, April 30 – May 2, 2009, Sparks, Nevada, USA*, SIAM, 733–744.
- [22] Wedel, M. and Kamakura, W. (2003). *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers, Norwell.
- [23] Younis, K.S. (1999). *Weighted Mahalanobis distance for hyper-ellipsoidal clustering*, Ph.D. thesis. Air Force Institute of Technology, Ohio.
- [24] Zakrzewska, D. (2007). On integratig unsupervised and supervised classification for credit risk evaluation. *Information technology and control*, 36, 98–102.