

Identification and Classification of the Best Communities That Ensure the Mastery of Their Expenditure by Using the Threshold of Their Cluster

Anouar Riad Solh¹ & Mourad EL Belkacemi²

¹Laboratoire Conception et Systèmes (Microélectronique et Informatique), Faculty of Sciences Rabat, University Mohammed V-Agdal, Rabat, Morocco

Correspondence: Anouar Riad Solh, Laboratoire Conception et Systèmes (Microélectronique et Informatique), Faculty of Sciences Rabat – University Mohammed V-Agdal, Rabat, Morocco. Tel: 7-212-3777-1834/35/38. Fax: 212-3777-4261. E-mail: anouarriadsolh@yahoo.fr/mourad_prof@yahoo.fr

Received: December 24, 2014 Accepted: January 4, 2015 Online Published: March 18, 2015

doi:10.5539/mas.v9n4p284 URL: <http://dx.doi.org/10.5539/mas.v9n4p284>

Abstract

In the context of the rationalization of expenditure of local communities, we have developed a technique for segmentation of local communities according to their revenue ratios by using the algorithms of data mining, and identifying in the first place, their weight to control expenditure in their cluster. The latter is characterized by a threshold dependent on the number of its elements. In a second level, we built a new reorganization in their classification in order to increase the weight and subsequently to ensure better control and good management of local communities spending.

Keywords: data mining, unsupervised classification, clustering algorithms, threshold, separation distance, categorical data, revenue ratios, Current Cash Flow (CCF), Local Community (LC)

1. Introduction

Today, forecast of future budget of governments requires reflection on the sustainability of public finances, facing among other things, a continuing global economic crisis (bankruptcy of Greece as an example). Some decisions are already taken by some countries of the European Union like France, Spain, Great Britain, Germany, etc. on budgetary discipline and fiscal consolidation to reduce deficits and debt levels of countries. However, rationalization of public expenditures will be effective at the national level unless it is established at the regional level (For local communities that may be regional, provincial, prefectural, urban towns and rural communities).

Taking this fact into consideration, and as part of the rationalization of local government spending, we have developed a new classification technique to compare communities' spending based on a fundamental aggregate for financial analysis community, which is the current cash flow (CCF). This margin is the excess of operating revenue over operating expenses. Low margin indicates a difficult management, and a negative margin means that the community living is beyond its means. However, the comparison of communities as ranked by their CCF values is relative and does not produce reliable results to financial analysts because of the diversity of their resources and their revenues. This is why we have thought to introduce segmentation techniques and clustering (unsupervised classification) to group communities that share the same levels of revenue ratios.

Then, for each community, we measured the distance between the value of the CCF and the value of the largest CCF cluster to which it belongs, proposing a threshold that is not exceeded by the elements of each cluster. The threshold calculation depends essentially on all the distances and the number of elements in a cluster. Depending on the value of this distance, we can determine the classification of this community in the cluster to inform about its weight-control spending. Indeed, a short distance is a significant weight and a significant distance results in a low weight.

The communities who share the same distance (the case for example of centers of severities of clusters which have a separation distance of zero), can be compared based on the value of the threshold of their cluster. Lower is the separation distance of a community and higher is the value of the threshold of its cluster, better is the weight control of expenditure.

2. Weaknesses of the Classical Classification

The current cash flow (CCF), which represents the excess of operating revenue over operating expenses [11], is intended to finance investment. This is the current savings of the community. Our goal is to find the technical means to rationalize the expenditure of local communities living in critical financial situations, taking into account their CCF and their recipes.

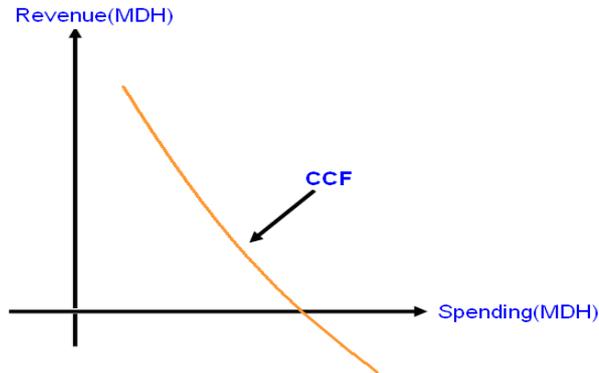


Figure 1. The shape of the CCF

Classification of communities according to the value of their CCF illustrates the control of their spending towards their income. In order to standardize and categorize the value of the CCF, we defined for each community LC_i the following report of its CCF categorization (NCCF):

$$NCCF(LC_i) = \frac{CCF(LC_i)}{MAX(CCF)}$$

where $CCF(LC_i)$ is the CCF value of community LC_i and $MAX(CCF)$ is the maximum value of CCF same type of communities (regions, provinces, prefectures, urban municipalities and rural communes). This ratio which is less than or equal to '1', tells us about the CCF value of LC_i over most CCF communities of the same type [1].

In a sample of 16 provinces in 2002, their report NCCF is shown [11] as follows:

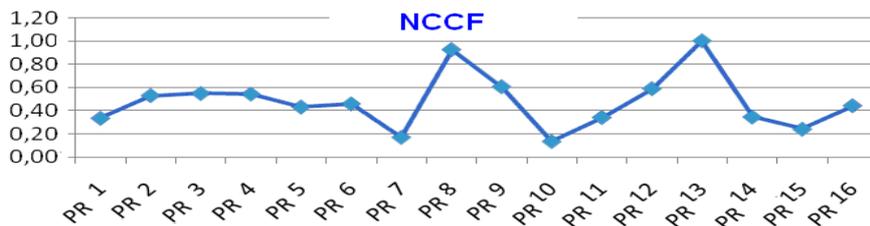


Figure 2. NCCF Report by province for year 2002

To make the correspondence between the NCCF report and the weight of expenditure control (category CCF) [2, 3], we proposed the following categorization:

Table 1. Correspondence between NCCF and Category of the CCF

Indicator	Operator	Margin	Category CCF	Situation of control spending
NMAC	<	0	Category 6	Critical state
	in	[0, 0.2[Category 5	very low
	in	[0.2, 0.4[Category 4	low
	in	[0.4, 0.6[Category 3	average
	in	[0.6, 0.8[Category 2	well
	>=	0.8	Category 1	very well

The more the CCF report categorization (NCCF) is closed to '1', better is the weight-control spending, and vice versa. If this ratio is far from '1', the weight control reflects a bad situation in critical expense.

Table 2. The result of the traditional classification of 16 provinces (previous example) according to CCF allows developing the following table

LC	NMAC in	Category CCF	Situation of control spending
PR 10	[0, 0.2[Category 5	Very low
PR 7	[0, 0.2[Category 5	Very low
PR 15	[0.2, 0.4[Category 4	Low
PR 1	[0.2, 0.4[Category 4	Low
PR 11	[0.2, 0.4[Category 4	Low
PR 14	[0.2, 0.4[Category 4	Low
PR 5	[0.4, 0.6[Category 3	Average
PR 16	[0.4, 0.6[Category 3	Average
PR 6	[0.4, 0.6[Category 3	Average
PR 2	[0.4, 0.6[Category 3	Average
PR 4	[0.4, 0.6[Category 3	Average
PR 3	[0.4, 0.6[Category 3	Average
PR 12	[0.4, 0.6[Category 3	Average
PR 9	[0.6, 0.8[Category 2	Well
PR 8	[0.8, 1]	Category 1	Verywell
PR 13	[0.8, 1]	Category 1	Verywell

In this case, the comparison of the CCF local communities regardless of their means (revenue) remains incomplete because of the diversity of their ways. So it is clear that this classification has deficiencies and weaknesses due to the different levels of revenue ratios: the share of value added tax in the revenue from taxation in revenue and gross savings rate....

The classification will be relevant when the LC have the same means and the same reference point (having the same center of gravity). That is why we introduce a segmentation technique (clustering) [4, 5] to group and classify LC who shares the same level of revenue ratios [11].

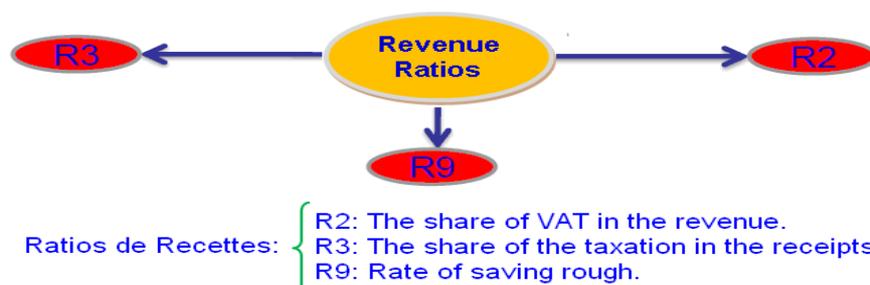


Figure 3. Revenue ratios

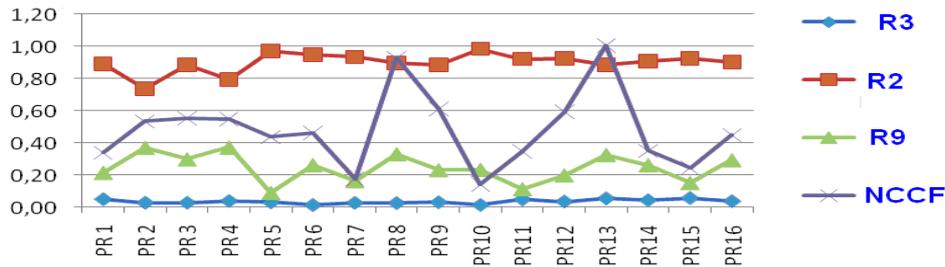


Figure 4. The following chart describes the layout of NCCF for a sample of 16 provinces (previous example) in relation to their revenue ratios for 2002

3. Setting the Points of Reference and Ensuring a Relevant Classification of LC according to Their Means

In order to make a relevant and accurate classification, we segmented first the LC [4,5] according to their local revenue ratios [11] by producing a set of clusters [6, 7, 8] (a cluster groups the LCs who share the same level of revenue ratios).

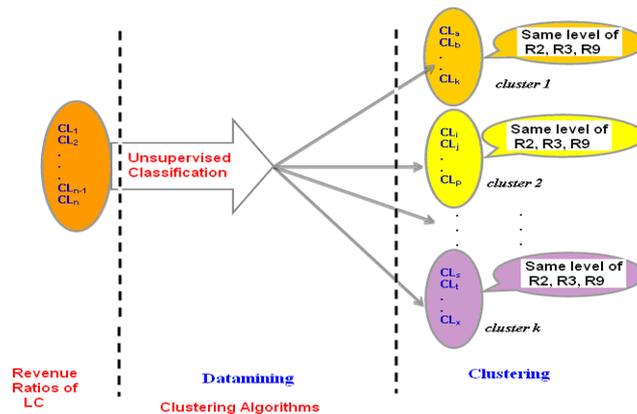


Figure 5. Segmentation of LC according to their revenue ratios by clustering algorithms

In our publication for the unsupervised classification (clustering) we used the tow-step algorithms, K-means and Kohonen [4, 5, 6, 7, and 8].

As a result, for each community, we measured the distance between [9, 10] the value of the CCF and the value of the largest CCF cluster to which it belongs. Knowing this distance we can determine the weight-control spending. A small distance corresponding to a significant weight and a large distance results in a low weight. This distance should not exceed a threshold [12, 13] which depends on the distances and the number of items in each cluster.

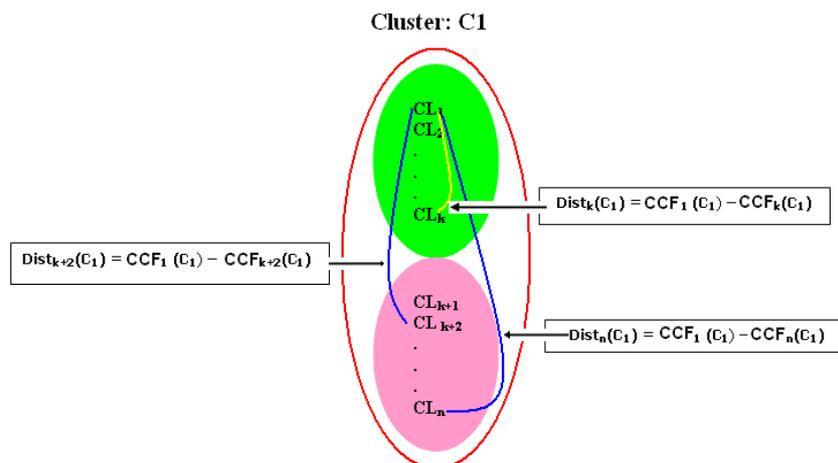


Figure 6. Calculating distances in a cluster

At the cluster level 'C₁', the community LC₁ has the largest CCF which is 'CCF₁' (center of gravity of the cluster 'C₁')

Dist_k(C_i) : It's the distance between the CCF community LC_k of cluster C₁ 'CCF_k(C₁)' and the highest CCF 'CCF₁(C₁)' of cluster C₁

4. Description of the Method

The weight of cost control of a community is measured by the distance between [9.10] the value of the CCF and the value of the best RTC cluster where it belongs.

Let K: IS THE NUMBER OF LOCAL A CLASSIFICATION (LC₁, LC₂, ...LC_K)

We refer to algorithms unsupervised classification (Two-step, K-means and Kohonen) [4, 5, 6, 7, 8] for the segmentation of communities according to their revenue ratios [11], producing a set of clusters. Each cluster includes the LC who shares the same level of revenue ratios.

Let M: IS THE NUMBER OF CLUSTERS (C₁, C₂, ..., C_M) RESULTS.

Let C_i : a cluster that includes LC n, we measure the high value of the RTC cluster C_i by: MAXCCF (C_i) with i <= M.

Let CCF_i(C_i) : the value of the CCF LC₁ the cluster C_i, we measure the distance between the value CCF_i(C_i) and the value MAXCCF(C_i) .

As the difference between these two values is close to zero the weight-control spending LCt is important, and vice versa. That is to say, the higher the value of the difference is far from zero, the community LCt suffers more from a mismanagement at its expenses.

5. The Threshold of a Cluster

Either C_k a cluster which includes a number n of local communities therefore:

$$n = \text{Card}(C_k)$$

Either MAXCCF(C_k): great CCF cluster C_k.

Either MINCCF(C_k) : Early CCF cluster C_k.

$$\text{Gdist}(C_k) = \text{MAXCCF}(C_k) - \text{MINCCF}(C_k)$$

This is the great separation distance for the cluster C_k having the distance between the high CCF and the small CCF the cluster C_k.

$$\text{Dist}_i(C_k) = \text{CCFMAX}(C_k) - \text{CCF}_i(C_k)$$

This value represents the distance separating the large CCF and the CCF of the community 'i' of the cluster C_k

We define the threshold S_k cluster C_k by the following formula [12, 13]:

$$S_k = \text{Gdist}(C_k) - \sum_{i=0}^n \frac{\text{Dist}_i(C_k)}{n}$$

In a cluster, the community with the best CCF represents the centre of gravity of the cluster (distance of separation equal '0').

As well, the centers of gravities of all the clusters have a separation distance equal to '0'.

In addition, a cluster that has a high threshold implies that the center of gravity is very far from the critical financial situation. And conversely a cluster that has a low threshold implies that the center of gravity is very close to the critical financial situation.

As a result: More separation distance of a local community is low and the threshold value of its cluster is higher the better the weight-control spending.

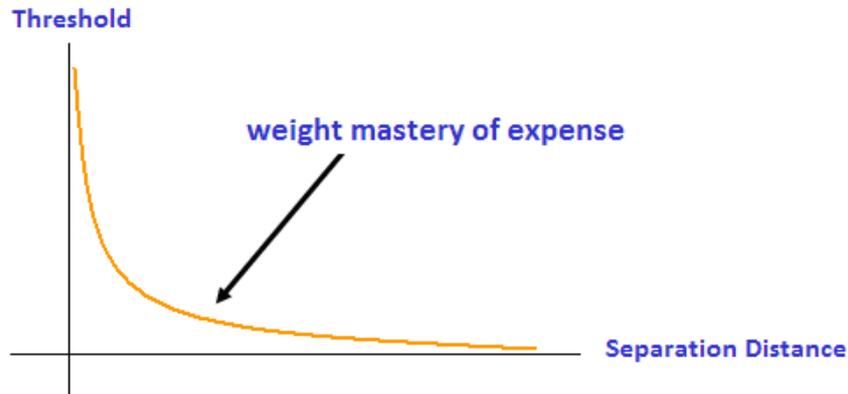


Figure 6. weight mastery of expense based on the threshold and the separation distance

To compare the weight-control spending centers of gravity of clusters, it is necessary to compare the thresholds of their cluster.

Either C_1 a cluster that has the threshold S_1 and which comprises a set of local communities that share the same levels of revenue ratios.

Either C_2 a cluster that has the threshold S_2 and which comprises a set of local communities that share the same levels of revenue ratios.

Either X the community which owns the largest CCF at cluster level C_1 (center of gravity of the cluster C_1) in this case the separation distance of the community X equal: '0'.

Either Y the community which owns the largest CCF at cluster level C_2 (center of gravity of the cluster C_2) in this case the separation distance of the community Y equal: '0'.

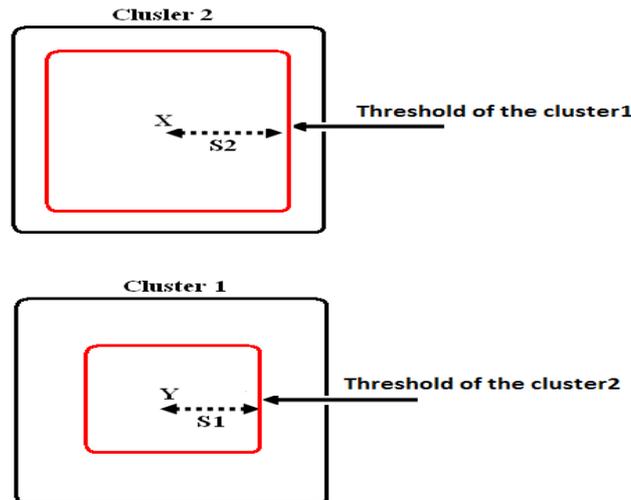


Figure 7. Threshold of cluster

To compare the weight of control of expenditure of these two communities X and Y, sharing the same distance of separation, we compare the thresholds of their cluster (S_1 and S_2).

$S_2 > S_1$ Implies that the community X is more remote from the critical financial situation that the community Y. In this case X has a weight of control of expenditure better than that of the community Y.

6. Evolution of CCF's Category of a Local Community

Let C be a cluster that includes n local communities.

Let X is the community that is the center of gravity of the cluster C and has the CCF category 'i'.

In order that the communities of cluster C provide a rationalization of their expenses, they must have the same type

of CCF than the X community.

In this case the rationalization of expenditures of these communities cluster C requires the migration of their category CCF to the center of gravity of the cluster C.

Let X, Y, Z and V communities in cluster C.

Community Y has the category CCF 'j'.

The community has Z category CCF 'k'.

The community has the V category CCF 'p'.

Rationalization of expenditure by local Y, Z and V requires the migration of their CCF category to category 'i'

CL \ Categorie	i	j	k	p
X	+			
Y	← +			
Z	← ← +			
V	← ← ← +			

Table 3. Table of evolution of the category CCF

- The rationalization of expenditure by the community Y, requires only a move of one step in its class of CCF.
- The rationalization of expenditure by the Z community requires a move of two steps in its class CCF.
- The rationalization of expenditure by the community V requires no changes in three-class CCF.

The more category CCF of a community is away from the center of gravity of the cluster, the more streamlined procedure applied by these leaders will be rigorous.

7. Some Results

Table 4. The following figure shows the necessary changes in the categories of the CCF for a sample of 16 provinces (LC) in order to ensure the rationalization of expenditures based on the (tow-step algorithm) segmentation of their financial data (revenue ratios and CCF) for 2002.

LC \ category	category1	category2	category3	category4	category5	category6
PR1				X		
PR2			X			
PR3	Y ←		Z			
PR4			X			
PR5		Y ←	Z			
PR6			X			
PR7		Y ←			Z	
PR8	X					
PR9		X				
PR10			Y ←		Z	
PR11				X		
PR12		Y ←	Z			
PR13	X					
PR14	Y ←				Z	
PR15				X		
PR16	Y ←		Z			

X: The community belongs to the same category of CCF that the center of gravity of the cluster (balanced management of expenses over revenue).

Z: Given these recipes, the community is poorly placed in this category because of the mismanagement of expenses over revenue.

Y: Given these recipes, the community needs this type of CCF to avoid difficult financial and critical situations and ensure proper rationalization of spending (the category of the center of gravity of the cluster where it belongs).

Table 5. The situation of local communities for Cluster1 (Threshold : 0,006764065 (B))

Local Community	NCCF	Category CCF	Separation Distance (A)	center of gravity	(B)- (A)
PR2	0,530322399	Category3	0,013528131	NO	-0,006764065
PR4	0,54385053	Category3	0	YES	0,006764065

Table 6. The situation of local communities for Cluster4 (Threshold=0,06503711 (B))

Local Community	NCCF	Category CCF	Separation Distance (A)	center of gravity	(B)- (A)
PR1	0,335899417	Category4	0,00796516	NO	0,05707195
PR11	0,34386458	Category4	0,00000000	YES	0,06503711
PR15	0,242326333	Category4	0,10153825	NO	-0,03650114

Table 7. The situation of local communities for Cluster2(Threshold=0,305516612 (B))

Local Community	NCCF	Category CCF	Separation Distance (A)	center of gravity	(B)- (A)
PR3	0,548654913	Category3	0,451345087	NO	-0,145828475
PR8	0,924506567	Category1	0,075493433	NO	0,230023179
PR13	1	Category1	0,00000000	YES	0,30551661
PR14	0,347432746	Category4	0,652567254	NO	-0,347050643
PR16	0,44415256	Category3	0,55584744	NO	-0,250330828

Table 8. The situation of local communities for Cluster5(Threshold=0,160792302 (B))

Local Community	NCCF	Category CCF	Separation Distance (A)	center of gravity	(B)- (A)
PR6	0,459577158	Category3	0,00000000	YES	0,160792302
PR10	0,137992555	Category5	0,321584603	NO	-0,160792302

Table 9. The situation of local communities for Cluster3 (Threshold=0,277432043 (B))

Local Community	NCCF	Category CCF	Separation Distance (A)	center of gravity	(B) – (A)
PR5	0,432759711	Category3	0,174039475	NO	0,103392569
PR7	0,173168505	Category5	0,433630681	NO	-0,433630681
PR9	0,606799185	Category2	0,00000000	YES	0,277432043
PR12	0,589674791	Category3	0,017124394	NO	0,260307649

The community which has a separation distance greater than the value of the threshold of its cluster [(B) – (A) < 0] indicates that this community does not ensure the mastery this its expenditures by report has its revenues.

Otherwise, more that this difference is negative more than the weight of control of its spending is low (example: for cluster 2, the community PR₃ has a very poor mastery of its expenditure)

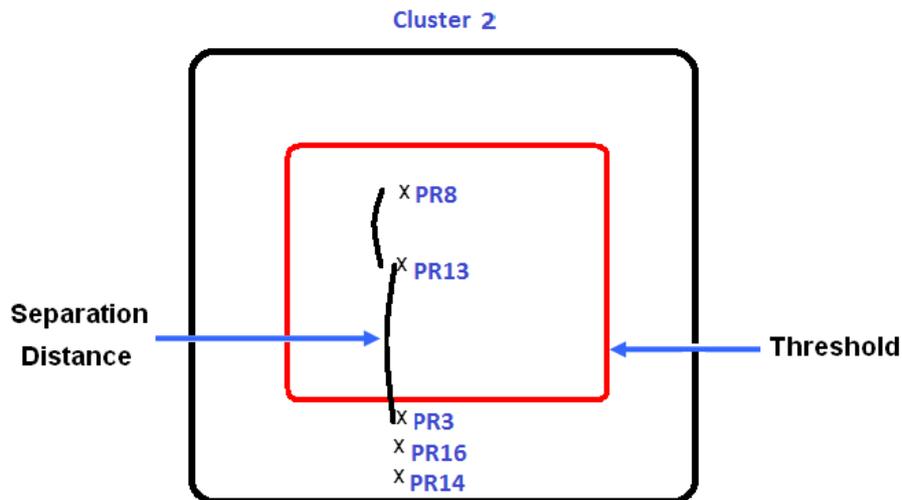


Figure 8. The situation of local communities for Cluster2

8. Conclusion

Rationalization of local government spending is a major project that requires real political will and the use of new technologies by using suitable algorithms to develop strategic dashboards and alerts to identify indicators. For this reason we have developed a new approach that presents an objective classification of local authorities in control of their spending over their means in order to increase their weight-control spending and ensure through following a better management of local government spending.

In addition we have presented a new method for a classification of the best communities that provide control of their expenditure on the basis of the value of the threshold of their cluster.

The classification of LCs within their means is an important step towards the rationalization of spending but is not sufficient.

References

- Bhatti, T. S., Bansal, R. C., & Kothari, D. P. (2002). Reactive Power Control of Isolated Hybrid Power Systems. *Proceedings of International Conference on Computer Application in Electrical Engineering Recent Advances (CERA)* (pp. 626-632). Indian Institute of Technology Roorkee (India).
- Chang, C. H., & Ding, Z. K. (2005). Categorical data visualization and clustering using subjective factors. *Data & Knowledge Engineering*, 53(2005), 243–262. <http://dx.doi.org/10.1016/j.datak.2004.09.001>
- Do, H. J., & Kim, J. Y. (2009). Clustering Categorical Data based on Combinations of Attribute Values. *International Journal of Innovative Computing, Information and Control*, 5(12(A)).
- Guide to financial analysis of the local communities retrospective approach. (2004). Retrieved from <http://www.mdemth.ma/sites/default/files/documents/files/56.pdf>
- Holmes, F. (2005). Comparison of Distance Measures in Cluster Analysis with Dichotomous Data. *Journal of Data Science*, 3(2005), 85-100.
- Ivancsy, R., & Kovacs, F. (2006). Clustering Techniques Utilized in Web Usage Mining. In *Proceedings of the 5th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases* (pp. 237- 242).
- Ming, Y. S., Jheng, J. W., & Lien, F. L. (2010). A Two-Step Method for Clustering Mixed Categorical and Numeric Data Tamkang. *Journal of Science and Engineering*, 13(1), 11-19.
- Mohamed, M. G. (2005). Shonali Krishnaswamy, Arkady Zaslavsky, Resource-aware Mining of Data Streams. *Journal of Universal Computer Science*, 11(8), 1440-1453.
- Nong, Y., & Li, X. Y. (2002). A scalable incremental learning algorithm for classification problems. *Computers & Industrial Engineering*, 43, 677–692. [http://dx.doi.org/10.1016/S0360-8352\(02\)00132-8](http://dx.doi.org/10.1016/S0360-8352(02)00132-8)
- Panagiotis, A., Christos, M., & Nikos, T. (2007). Algorithms for Clustering ClickStream Data, Preprint submitted to *Information Processing Letters - IPL*.

- Shaeela, A., Tasleem, M., Ahsan, R. S., & Khan, M. I. (2010). Data Mining Model for Higher Education System. *European Journal of Scientific Research*, 43(1), 24-29.
- Sivaram, N., & Ramar, K. (2010). Applicability of Clustering and Classification Algorithms for Recruitment Data Mining. *International Journal of Computer Applications (0975 – 8887)*, 4(5). <http://dx.doi.org/10.5120/823-1165>
- Smruti, R. S., & Karin, M. (2010). DUST: A Generalized Notion of Similarity between Uncertain Time Series KDD10 Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, DC, USA.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).