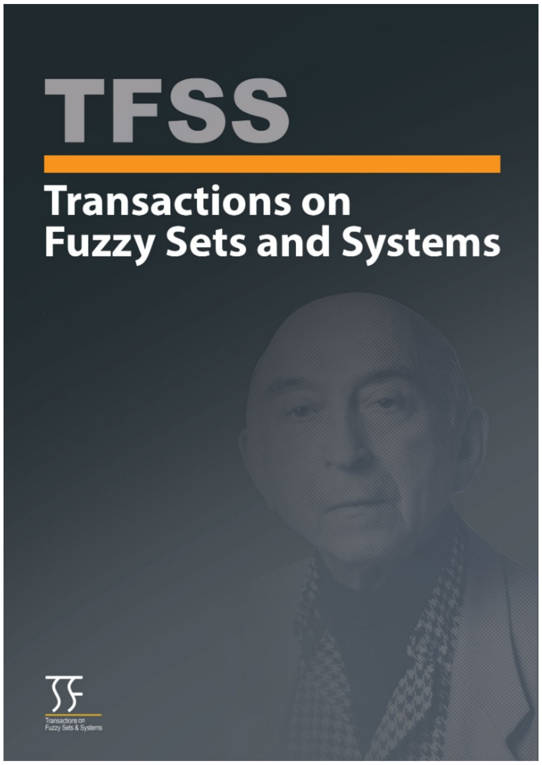
Transactions on Fuzzy Sets and Systems



Transactions on Fuzzy Sets and Systems

ISSN: **2821-0131**

https://sanad.iau.ir/journal/tfss/

A New Approach to Define the Number of Clusters for Partitional Clustering Algorithms

Vol.3, No.1, (2024), 67-87. DOI: http://doi.org/10.30495/tfss.2023.1990425.1078

Author(s):

Huliane Medeiros da Silva, Departamento de Engenharias e Tecnologias, Universidade Federal Rural do Semi-Árido, Pau dos Ferros, Brasil. E-mail: huliane@ufersa.edu.br
Benjamín René Callejas Bedregal, Departamento de Informática e Matemática Aplicada, Universidade Federal do Rio Grande do Norte, Natal, Brasil. E-mail: bedregal@dimap.ufrn.br
Anne Magály de Paula Canuto, Departamento de Informática e Matemática Aplicada, Universidade Federal do Rio Grande do Norte, Natal, Brasil. E-mail: anne.canuto@ufrn.br
Thiago Vinícius Vieira Batista, Departamento de Informática e Matemática Aplicada, Universidade Federal do Rio Grande do Norte, Natal, Brasil. E-mail: thiagovvb@gmail.com
Ronildo Pinheiro de Araújo Moura, Departamento de Informática e Matemática Aplicada, Universidade Federal do Rio Grande do Norte, Natal, Brasil. E-mail: ronildoamoura@gmail.com

Transactions on Fuzzy Sets and Systems (TFSS)
URL: https://sanad.iau.ir/journal/tfss/
Online ISSN: 2821-0131
Vol.3, No.1, (2024), 67-87
DOI: http://doi.org/10.30495/tfss.2023.1990425.1078

Article Type: Original Research Article

A New Approach to Define the Number of Clusters for Partitional Clustering Algorithms

Huliane Medeiros da Silva*, **Benjamín René Callejas Bedregal**, **Anne Magály de Paula Canuto**, **Thiago Vinícius Vieira Batista**, **Ronildo Pinheiro de Araújo Moura**

Abstract. Data clustering consists of grouping similar objects according to some characteristic. In the literature, there are several clustering algorithms, among which stands out the Fuzzy C-Means (FCM), one of the most discussed algorithms, being used in different applications. Although it is a simple and easy to manipulate clustering method, the FCM requires as its initial parameter the number of clusters. Usually, this information is unknown, beforehand and this becomes a relevant problem in the data cluster analysis process. In this context, this work proposes a new methodology to determine the number of clusters of partitional algorithms, using subsets of the original data in order to define the number of clusters. This new methodology, is intended to reduce the side effects of the cluster definition phase, possibly making the processing time faster and decreasing the computational cost. To evaluate the proposed methodology, different cluster validation indices will be used to evaluate the quality of the clusters obtained by the FCM algorithms and some of its variants, when applied to different databases. Through the empirical analysis, we can conclude that the results obtained in this article are promising, both from an experimental point of view and from a statistical point of view.

AMS Subject Classification 2020: 68T01; 68T05; 68T27
Keywords and Phrases: Partitional clustering algorithms, Clustering fuzzy, Number of cluster.

1 Introduction

The concept of data clustering consists of clustering similar objects together into groups, which are called clusters, taking into account one or more common features [1, 2]. In this context, several clustering techniques have been proposed in the literature, including hierarchical and partitional algorithms, which are widely used in several applications of different areas of knowledge. Most clustering algorithms require the number of clusters in the partition as an input parameter. However, the ideal number of clusters that represents the dataset, most of the time, is unknown. Thus, the definition of the number of clusters is one of the fundamental problems in the process of clustering data [3].

In the literature, several approaches have been proposed to determine the number of clusters [4, 5, 6, 7, 8, 9]. In spite of the potential of these approaches, there is no universal approach that performs well for all clustering problems [10]. The definition of the number of clusters in the data clustering process can take time and have a high computational cost, especially when a dataset has a large number of instances and attributes. In general, it is necessary to run an algorithm several times with different cluster settings, to

*Corresponding Author: Huliane Medeiros da Silva, Email: huliane@ufersa.edu.br, ORCID: 0009-0000-1420-3376
Received: 2 July 2023; **Revised**: 2 September 2023; **Accepted**: 3 September 2023; **Available Online**: 24 January 2024; **Published Online**: 7 May 2024.

How to cite: Silva HMd, Bedregal BR, Canuto AMP, Batista TVV, Moura RPA. A new approach to define the number of clusters for partitional clustering algorithms. *Transactions on Fuzzy Sets and Systems.* 2024; 3(1): 67-87. DOI: http://doi.org/10.30495/tfss.2023.1990425.1078

evaluate the result of each run, and then to select the best number of clusters. In addition, most applications have large datasets, which demand computational time and resources for processing. Therefore, the size of the dataset can become an aggravating factor in the definition of the number of clusters and, consequently, in the clustering process.

Despite presenting satisfactory results, the different approaches proposed in the literature to define the number of clusters [4, 5, 7, 8, 11, 12, 13, 14, 15] use the complete dataset, which can lead to problems such as high processing time and high computational cost. Consequently, the use of large datasets with high dimensions limits the application of a clustering algorithm, which can lead to an inefficient clustering process. Therefore, the main motivation of this work started from the need to try to improve the definition of the number of clusters of partitional algorithms, by proposing a new methodology to determine the number of clusters. The idea is to use a subset of the original data to define the number of clusters, in an attempt to mitigate the side effects of the cluster definition phase, possibly making processing time faster and lowering the computational cost.

2 The Proposed Approach

As previously mentioned, the phase of finding the best number of clusters for a dataset demands time and has a high computational cost, since it is performed by running an algorithm with different cluster configurations, evaluating each execution, and then evaluating each result to select the best number of clusters. In addition, the use of large datasets is an aggravating factor at this stage, due to the time and computational resources that this processing demands. In order to accelerate the definition of the number of clusters in the clustering process, this work presents a proposal for data partitioning applied to clustering algorithms, based on validation indices to determine the ideal subset of a dataset. The main objective of this proposal is to find, for each clustering algorithm considered in this research, a percentage p of the data that, for any dataset X we can choose a $p\%$ of the data in X to determine the best number of clusters. Our goal is to define an expressive p value in a way that its performance is similar to when the whole dataset is considered.

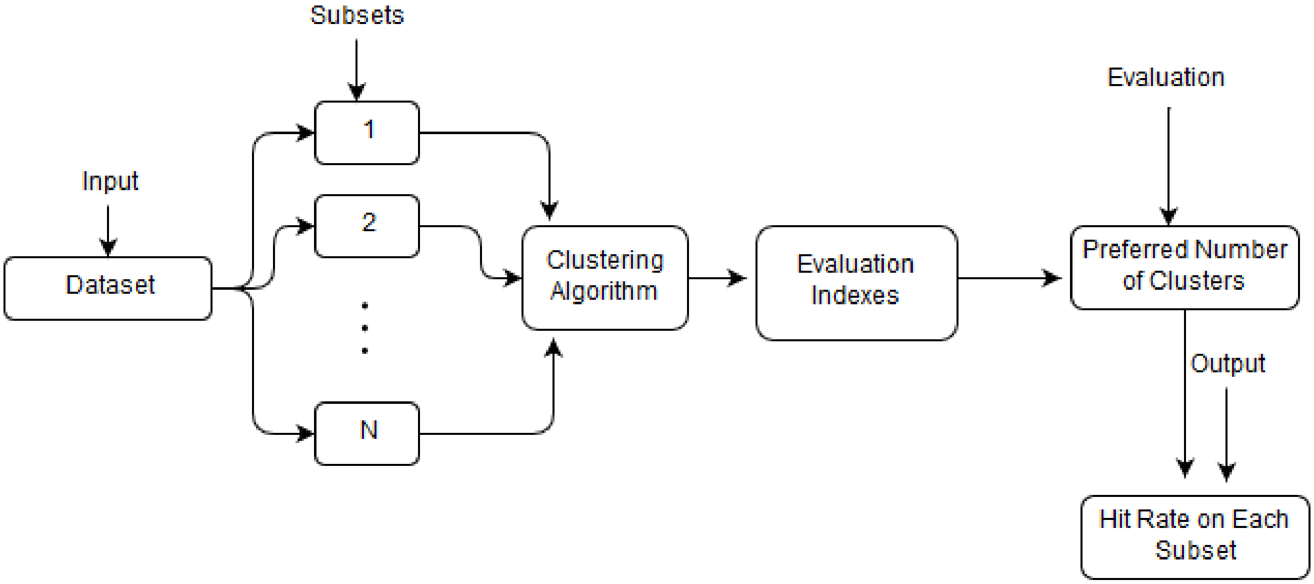


Figure 1: Structure of the proposed approach

Figure 1 presents the steps of the approach proposed in this work. A dataset is partitioned into N subsets.

Each subset has different sizes in relation to the total set, i.e., each subset has a particular percentage of the total set (p). The instances of each subset are randomly selected. and each subset has different sizes (with a variation of 10% each). The clustering algorithm is applied on all subsets and also on the full dataset ($p = 100\%$). Each clustering is evaluated based on validation indices and the best number of clusters for each index is defined. Finally, an analysis is performed on the number of clusters selected by each index, in an attempt to identify the minimum percentage of data sufficient to infer the number of clusters.

As an illustrative example, consider a given dataset with 1000 instances and 5 attributes. We need to cluster the instances in k clusters that best represent this dataset. Initially, the dataset is divided into N subsets, each of which has a size corresponding to a different percentage of the full dataset. Usually, we start with a low percentage. Then, there is a gradual increase of this percentage in the following subsets, until it reaches 100% ($p = 100\%$). For example, we can start with 10% of the original data and increase the percentage by 10%, until it reaches 100% of the data. In this case, we will have 10 subsets with 10, 20, ..., 100% of the original dataset. The next step is to apply clustering algorithms on each subset, varying the number of clusters k, $k_{min} \leq k \leq k_{max}$. Suppose we use $k_{min} = 2$ and $k_{max} = 10$. In this case, each clustering algorithm is applied to each subset of the original dataset, varying the number of clusters from 2 to 10. Then, the validation stage of the clusters is performed based on validation indices, in order to identify the subset that best represents the entire dataset. Each subset has the best cluster number defined according to the used validation index. For example, suppose that the best cluster number is $k = 3$ for $p = 10\%$ and $p = 100\%$ and $k = 4$ for $p = 50\%$, according to an index I. Note that the number of clusters set for 10% and 100% of the data is the same. In this example, the use of 10% of the data would be sufficient to infer the number of clusters that best represent the dataset. Finally, it is possible to identify the minimum data subset sufficient to infer the number of clusters that best represent the dataset, after analyzing the number of clusters defined by an index I in each subset.

3 Experimental Setting up

In order to evaluate the feasibility of the proposed method, an empirical analysis will be conducted. This section describes the main aspects of this empirical analysis.

3.1 Methods and Material

As previously discussed, the methodology proposed in this paper aims to investigate the minimum amount of data sufficient to infer the number of clusters in a dataset using partitional algorithms.

Initially, each dataset was normalized and partitioned into subsets of different sizes. The experiments carried out compared the performance of each subset in relation to the original dataset, through the performance obtained by the clustering validation indices (described in 3.4). The quality of a cluster is related to the ability to find the ideal number of clusters.

In order to evaluate the quality of the obtained partitions, an experimental methodology was proposed in order to investigate and identify the best number of clusters for the dataset. Each clustering algorithm runs on each subset and on the original dataset for each number of clusters k, $k_{min} \leq k \leq k_{max}$, aiming to find the best number of clusters in each subset as well as in the original dataset. Then, for each validation index, the number of clusters selected is stored. This experiment is repeated 31 times, with different initialization in each run [16]. The purpose of performing different initializations is to get as close as possible to the most likely number of clusters for a given dataset. The number of clusters is obtained after checking the most frequent value resulting from the 31 iterations. The variation in the number of clusters, k, being $k_{min} = 2$ and $k_{max} = 10$, took place based on the number of classes of all analyzed datasets.

3.2 Datasets

In this analysis, 30 datasets were imported from UCI Machine Learning Repository [17], Kaggle Datasets [18] and GitHub. The main features of these datasets are described in Table 1. All datasets had been preprocessed, with the goal of correcting some issues, such as attributes in different scales and missing values. In addition, it is noteworthy that they are supervised classification datasets and, for this reason, the class attribute was removed for the clustering process.

Table 1: Dataset features

Datasets	**Instances**	**Attributes**	**Classes**
Balance Scale	625	4	3
Banknote Authentication (BA)	1372	4	2
Bupa	345	6	2
Climate Model Simulation Crashes CMSC)	540	20	2
Cnae-9	1080	856	9
Column 3C	310	6	3
Contraceptive	1473	9	3
Ecoli	336	7	8
Glass Identification (GI)	214	9	6
Heart Statlog	270	13	2
Haberman	306	3	2
Ionosphere	351	34	2
Iris	150	4	3
Lymphography	148	18	4
Molecular Biology (MB)	3190	60	3
Multiple Features (MF)	2000	64	10
Parkinsons	195	22	2
Pima	768	8	2
Planning Relax	182	12	2
QSAR Biodegradation (QSAR-B)	1055	41	2
Robot Failure lp4	117	90	3
Seeds	210	7	3
Semeion	1593	256	10
Sonar	208	60	2
Steel Plates Faults (SPF)	1941	33	2
Thoracic Surgery (TS)	470	16	2
User Knowledge Modeling (UKM)	403	5	5
Vehicle	846	18	4
Voice	3168	20	2
Wine	178	13	3

3.3 Clustering Algorithms

Partitional algorithms can be divided into two approaches: *crisp* and *fuzzy*. In the *crisp* approach, each observation in the dataset belongs exclusively to a single cluster, while in the *fuzzy* approach, each object can belong to more than one cluster with a degree of relevance $u_{ij} \in [0,1]$. In this article, the focus is on the *fuzzy* approach, specifically the following algorithms: FCM [19, 20], ckMeans [21] and FCMσ [22].

Based on the concept of fuzzy logic, the Fuzzy C-means algorithm [19, 20] divides a dataset $X = \{x_1, x_2, ..., x_n\}$ into k clusters, resulting in a fuzzy partition matrix $[\mu_{ij}]_{n \times k}$, called a membership matrix [23].

The Fuzzy C-Means algorithm searches for the fuzzy partition of the dataset, which minimizes the objective function, given in Equation (1).

$$J = \sum_{i=1}^{n} \sum_{j=1}^{k} \mu_{ij}^{m} d(x_i; c_j)^2 \tag{1}$$

where m is the fuzzification parameter, [2] which defines the allowed distance between an object (point) and cluster centers; x_i is i-th data object; c_j is the center of the j-th cluster; μ_{ij} is the membership degree of x_i to j-th cluster; $d(x_i; c_j)$ is the distance between x_i and c_j. Note that c_j does not necessarily belong to the dataset, but has similar composition (same attributes) of the elements in the dataset.

The Fuzzy C-Means algorithm receives as input a dataset $X = \{x_1, x_2, ..., x_n\}$, the number of clusters k and the value of m. Then, it initializes the membership matrix μ; calculating the initial fuzzy membership matrix μ according to Equation (2).

$$\mu_{ij} = \frac{1}{\sum_{l=1}^{k} \left(\frac{d(x_i;c_j)}{d(x_i;c_l)} \right)^{\frac{2}{m-1}}} \tag{2}$$

After **calculating** J using Equation (1); the center of the cluster j is calculated using Equation (3).

$$c_j = \frac{\sum_{i=1}^{n} \mu_{ij}^{m} x_i}{\sum_{i=1}^{n} \mu_{ij}^{m}} \tag{3}$$

The algorithm continues to update of the fuzzy membership matrix according to Equation (2), as well as the centers of the clusters, according to Equation (3). This process continues until a stopping condition is reached. The two most usual stop conditions of clustering algorithms are: defining a fixed number of iterations and defining a threshold $\varepsilon > 0$, stopping the process when $||J^t - J^{(t-1)}|| \leq \varepsilon$, where J^t is the objective function calculated in the current iteration and $J^{(t-1)}$ is the objective function calculated in the previous iteration.

The other clustering algorithms considered in this research are variants from the FCM. The ckMeans algorithm follows the same FCM framework, however, it differs in how to calculate the cluster centers [24, 25]. Finally, the FCMσ changes the distance metric used in the conventional FCM to a new metric, taking into account a distance variation in each cluster [22].

3.4 Validation Indices

In clustering tasks, it is important to evaluate the resulting partition to determine the quality of the obtained solution, as well as whether it is satisfactory for the desired goal. The evaluation of the partition can be done by clustering validation indices, which can be broadly divided into two main categories: internal and external. The internal indices measure the similarity of a partition using the instances of the clusters obtained by the clustering algorithm. External indices use external information related to the partition, usually the class labels to evaluate the obtained partition [26, 27].

[2] The value of m influences directly on the resulting cluster. High values of m result in less well defined clusters [28].

In this article, three cluster validation indices will be used: (i) Modified Partition Coefficient (MPC); (ii) Sugeno; e (iii) Xie and Beni (XB). The first index involves only the membership matrix, while the second and third ones use membership matrix and information about the dataset. These are well-known indices and the formal definitions of these validation indices can be found easily in the literature[29, 30, 31].

4 Experimental Results

Tables 2, 3 and 4 illustrate the results obtained for all analyzed datasets, using the FCM algorithm and the MPC, XB and Sugeno indices, respectively. This experiment was replicated for the other clustering algorithms and validation indices. Tables 5, 6 and 7 present the results obtained for all analyzed datasets, using the ckMeans algorithm and the MPC, XB and Sugeno indices, respectively. Tables 8, 9 and 10 present the results obtained for all datasets analyzed, using the FCMσ algorithm and the MPC, XB and Sugeno indices, respectively. The rows of each table correspond to each dataset and the columns correspond to the percentage of the data of the corresponding subset. The values presented in each cell correspond to the number of groups selected for each dataset and the number of times this number was selected (in parentheses), out of 31 repetitions. Finally, the shaded cells indicate the cases in which the algorithm-index configuration selected the same number of groups in all analyzed percentages and in the original dataset.

Table 2: FCM-MPC: Number of clusters selected for multiple subsets of data.

Datasets	**Percentages of Data**									
	10%	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Balance	10(10)	4(9)	6(7)	4(6)	6(13)	6(18)	6(15)	6(25)	6(25)	6(30)
BA	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Bupa	2(23)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
CMSC	2(19)	2(28)	2(29)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Cnae-9	2(5)3(5)	3(5)5(5)8(5)	3(9)	3(9)	9(8)	2(9)	10(6)	2(7)	2(6)	3(10)
Column 3C	2(28)	2(29)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Contraceptive	2(25)	2(27)	2(28)	2(30)	2(30)	2(30)	2(31)	2(31)	2(31)	2(31)
Ecoli	10(7)	3(24)	3(22)	3(27)	3(28)	3(29)	3(31)	3(30)	3(31)	3(31)
GI	2(20)	2(24)	2(27)	2(21)	2(17)	2(19)	3(18)	2(16)	3(19)	3(23)
Haberman	10(15)	3(7)	4(9)	4(14)	4(16)	4(19)	4(16)	3(14)4(14)	4(15)	3(19)
Heart Statlog	2(11)	2(22)	2(28)	2(27)	2(31)	2(31)	2(31)	2(31)	2(30)	2(31)
Ionosphere	2(15)	2(30)	2(29)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Iris	2(29)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Lymph	10(31)	10(20)	2(18)	2(14)	2(15)	2(16)	2(16)	2(18)	2(19)	2(28)
MB	10(23)	10(15)	10(21)	10(13)	10(11)	10(9)	10(12)	9(8)	10(8)	10(11)
MF	3(16)	2(21)	2(17)	2(23)	2(22)	2(20)	2(24)	2(24)	2(20)	2(23)
Parkinsons	2(14)	3(19)	3(23)	3(25)	3(28)	3(31)	3(30)	3(30)	3(27)	3(28)
Pima	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Planning Relax	10(31)	10(26)	2(12)	3(11)	2(18)	2(15)	3(14)	3(12)	3(12)	2(22)
QSAR-B	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Robot Failure	10(31)	10(28)	2(23)	2(24)	2(27)	2(28)	2(29)	2(28)	2(29)	2(26)
Seeds	10(15)	2(28)	2(30)	2(30)	2(31)	2(31)	2(30)	2(31)	2(30)	2(31)
Semeion	3(9)	2(8)	2(10)	2(13)	2(11)	2(18)	2(12)	2(19)	2(16)	2(18)
Sonar	10(29)	2(9)	2(19)	2(18)	2(23)	2(18)	2(20)	2(15)3(15)	2(19)	2(27)
SPF	2(18)	2(25)	2(29)	2(30)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)
TS	2(26)	2(23)	2(31)	2(30)	2(30)	2(31)	2(31)	2(29)	2(31)	2(31)
UKM	10(21)	10(7)	2(17)	2(21)	2(21)	2(26)	2(29)	2(30)	2(31)	2(29)
Vehicle	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Voice	2(27)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Wine	10(12)	4(11)	3(12)	4(13)	3(13)4(13)	3(12)4(12)	4(15)	4(17)	4(20)	4(19)

Table 3: FCM-XB: Number of clusters selected for multiple subsets of data

Datasets	**Percentages of Data**									
	10%	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Balance	10(11)	9(11)10(11)	10(12)	10(17)	10(13)	10(17)	10(22)	10(22)	10(24)	10(18)
GI	2(7)	2(12)	2(11)	2(9)5(9)6(9)	2(11)	2(11)	2(9)	5(18)	2(12)	5(30)
Bupa	2(10)	2(19)	2(16)	2(25)	2(26)	2(23)	2(28)	2(28)	2(29)	2(31)
CMSC	2(23)	2(29)	2(29)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Cnae-9	2(8)3(8)	2(13)	2(13)	2(9)3(9)	2(14)	2(13)	2(15)	2(16)	2(16)	2(13)
Column 3C	10(13)	2(24)	2(28)	2(28)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Contraceptive	2(8)	3(20)	3(26)	3(20)	3(22)	3(24)	3(29)	3(26)	3(27)	3(30)
Ecoli	10(7)	9(7)	3(9)	3(12)	3(8)	3(14)	3(12)	3(14)	3(9)9(9)	8(12)
GI	2(15)	2(17)	2(25)	2(21)	2(17)	2(18)	3(19)	2(16)	3(19)	3(23)
Haberman	3(6)8(6)10(6)	3(10)	8(7)	3(13)	3(14)	7(8)	3(12)	3(13)	3(11)	3(18)
Heart Statlog	10(8)	9(6)10(6)	5(9)	7(6)8(6)9(6)10(6)	10(10)	9(7)10(7)	10(9)	8(11)	9(9)	9(14)
Ionosphere	2(16)	2(25)	2(25)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Iris	2(28)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Lymph	10(29)	2(9)	2(17)	3(16)	3(17)	3(19)	3(16)	3(19)	3(17)	3(18)
MB	10(26)	10(25)	10(23)	10(26)	10(20)	10(28)	10(22)	10(22)	10(21)	10(23)
MF	2(12)	2(20)	2(15)	2(21)	2(22)	2(13)	2(22)	2(21)	2(15)	2(25)
Parkinsons	2(16)	3(17)	3(24)	3(24)	3(28)	3(31)	3(31)	3(30)	3(30)	3(29)
Pima	2(28)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Planning Relax	10(20)	10(9)	2(14)	2(11)	2(16)	2(16)	2(12)	2(12)	2(18)	2(21)
QSAR-B	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Robot Failure	10(31)	10(17)	2(21)	2(25)	2(26)	2(26)	2(29)	2(28)	2(29)	2(26)
Seeds	10(9)	2(7)	5(5)7(5)	7(9)	6(8)	6(6)7(6)10(6)	6(9)	9(9)	6(8)9(8)	10(14)
Semeion	10(9)	10(10)	10(7)	9(8)	10(12)	10(7)	10(7)	10(7)	10(10)	10(9)
Sonar	10(18)	3(15)	2(17)	3(14)	2(17)	3(16)	3(15)	3(20)	3(22)	3(25)
SPF	2(23)	2(28)	2(25)	2(29)	2(31)	2(30)	2(31)	2(31)	2(31)	2(31)
TS	2(26)	2(24)	2(31)	2(30)	2(30)	2(31)	2(31)	2(29)	2(31)	2(31)
UKM	9(11)	9(10)10(10)	10(15)	9(12)	10(9)	7(8)	10(8)	6(8)	7(15)	7(13)
Vehicle	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Voice	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Wine	3(11)	3(13)	4(13)	3(17)	3(16)	3(15)4(15)	4(16)	4(22)	4(19)	4(19)

Table 4: FCM-Sugeno: Number of clusters selected for multiple subsets of data

Datasets	**Percentages of Data**									
	10%	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Balance	10(22)	10(19)	10(25)	10(27)	10(25)	10(26)	10(29)	10(26)	10(29)	10(27)
GI	10(11)	9(10)10(10)	10(11)	9(15)	9(16)	9(19)	9(19)	9(27)	9(16)	9(17)
Bupa	10(13)	6(6)	10(7)	5(7)	5(6)	5(9)	7(8)	5(12)	5(10)7(10)	5(21)
CMSC	8(8)	9(9)	4(10)	5(8)	5(15)	5(10)	10(7)	5(9)	5(10)	5(18)
Cnae-9	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Column 3C	10(23)	10(10)	4(10)	5(10)	5(16)	5(14)	5(14)	4(15)	5(23)	5(20)
Contraceptive	10(19)	10(24)	10(28)	10(28)	10(28)	10(29)	10(29)	10(30)	10(31)	10(31)
Ecoli	10(24)	10(17)	10(17)	10(10)	10(10)	9(10)	10(11)	10(11)	7(13)	7(14)
GI	10(13)	3(10)	4(14)	3(10)	4(14)	3(13)	3(18)	3(15)	3(19)	3(23)
Haberman	10(12)	10(14)	10(10)	10(9)	10(13)	10(12)	10(13)	10(14)	10(22)	10(20)
Heart Statlog	10(17)	10(15)	10(16)	9(16)	10(15)	10(15)	9(11)	8(13)	10(14)	9(10)
Ionosphere	10(22)	10(20)	10(15)	10(19)	10(18)	10(16)	10(22)	10(20)	10(23)	10(20)
Iris	10(15)	2(29)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Lymph	10(31)	10(29)	10(29)	10(26)	10(28)	10(30)	10(31)	10(31)	10(30)	10(31)
MB	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
MF	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Parkinsons	10(8)	3(19)	3(23)	3(25)	3(28)	3(31)	3(31)	3(30)	3(29)	3(29)
Pima	8(6)	3(10)	2(7)	3(9)	3(14)	3(15)	3(14)	3(22)	3(22)	3(28)
Planning Relax	10(31)	10(30)	10(30)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
QSAR-B	5(8)	5(16)	6(11)	5(15)	5(19)	6(16)	5(16)	5(18)	5(21)	5(30)
Robot Failure	10(31)	10(31)	10(29)	10(27)	10(25)	10(29)	10(31)	10(31)	10(31)	10(31)
Seeds	10(23)	10(15)	7(8)9(8)10(8)	10(12)	10(11)	9(12)	9(9)10(9)	9(11)	9(10)10(10)	10(14)
Semeion	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Sonar	10(31)	10(25)	10(26)	10(29)	10(29)	10(29)	10(30)	10(30)	10(31)	10(31)
SPF	3(9)	5(11)	5(16)	5(17)	5(19)	5(18)	5(20)	7(16)	7(19)	7(27)
TS	2(21)	2(19)	2(29)	2(29)	2(30)	2(31)	2(31)	2(29)	2(31)	2(31)
UKM	10(27)	10(22)	10(20)	10(21)	10(25)	10(29)	10(26)	10(25)	10(27)	10(25)
Vehicle	4(12)	2(18)	2(15)	2(16)	2(21)	2(22)	2(19)	2(21)	2(27)	2(31)
Voice	4(13)	4(13)	4(19)	4(16)	4(25)	4(21)	4(21)	4(22)	4(21)	4(31)
Wine	10(15)	10(10)	4(9)	4(10)	4(12)	4(13)	4(15)	4(19)	4(18)	4(19)

Table 5: ckMeans-MPC: Number of clusters selected for multiple subsets of data

Datasets	**Percentages of Data**									
	10%	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Balance	4(8)	2(8)	4(12)	4(13)	3(14)	4(12)	4(17)	4(14)	3(15)	3(13)4(13)
GI	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Bupa	2(22)	2(29)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
CMSC	2(24)	2(30)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Cnae-9	3(12)	3(11)	3(18)	3(22)	3(22)	3(17)	3(22)	3(23)	3(22)	3(23)
Column 3C	2(30)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Contraceptive	2(29)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(30)	2(31)
Ecoli	3(15)	3(20)	3(24)	3(20)	3(24)	3(23)	3(24)	3(29)	3(30)	3(30)
GI	2(18)	2(28)	2(29)	2(29)	2(31)	2(31)	2(31)	2(30)	2(30)	2(31)
Haberman	10(7)	3(13)	3(13)	3(19)	4(16)	3(17)	3(21)	3(25)	3(28)	3(20)
Heart Statlog	2(13)	2(23)	2(25)	2(27)	2(29)	2(29)	2(31)	2(31)	2(31)	2(31)
Ionosphere	2(25)	2(27)	2(31)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Iris	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Lymph	10(30)	10(11)	2(25)	2(21)	3(18)	2(18)	3(23)	3(21)	3(21)	3(29)
MB	2(22)	2(27)	2(28)	2(30)	2(31)	2(31)	2(30)	2(31)	2(30)	2(30)
MF	2(24)	2(26)	2(28)	2(30)	2(31)	2(31)	2(29)	2(31)	2(31)	2(31)
Parkinsons	2(15)	2(19)	2(23)	2(24)	2(24)	2(21)	2(28)	2(28)	2(30)	2(31)
Pima	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Planning Relax	10(26)	10(15)	2(11)	2(14)	3(11)	2(16)	3(15)	2(16)	2(21)	2(27)
QSAR-B	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Robot Failure	10(31)	10(19)	2(17)	2(28)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)
Seeds	10(11)	2(24)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Semeion	2(14)	2(20)	2(16)	2(22)	2(24)	2(20)	2(19)	3(19)	3(18)	3(24)
Sonar	10(21)	2(15)	2(15)3(15)	2(15)3(15)	2(17)	2(17)	2(18)	2(19)	2(16)	2(29)
SPF	2(23)	2(26)	2(29)	2(31)	2(30)	2(30)	2(31)	2(31)	2(31)	2(31)
TS	2(31)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
UKM	10(8)	2(18)	2(27)	2(22)	2(23)	2(29)	2(27)	2(28)	2(29)	2(31)
Vehicle	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Voice	2(26)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Wine	10(14)	4(12)	3(15)	3(15)	3(17)	3(18)	3(13)	3(14)	3(16)	4(30)

Table 6: ckMeans-XB: Number of clusters selected for multiple subsets of data.

Datasets	**Percentages of Data**									
	10%	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Balance	10(10)	10(16)	10(21)	10(22)	10(26)	10(24)	10(26)	10(23)	10(27)	10(27)
GI	8(10)	9(11)	10(8)	9(11)	10(10)	9(12)	10(10)	10(12)	10(12)	10(13)
Bupa	10(13)	2(10)	10(12)	10(11)	10(14)	10(15)	10(14)	10(11)	10(10)	10(8)
CMSC	2(28)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Cnae-9	10(27)	10(25)	10(27)	10(24)	10(24)	10(23)	10(26)	10(23)	10(22)	10(26)
Column 3C	10(13)	10(16)	10(14)	10(11)	10(21)	10(19)	10(20)	10(19)	10(21)	10(21)
Contraceptive	10(15)	10(17)	10(21)	10(18)	10(20)	10(20)	10(16)	9(14)10(14)	10(19)	10(17)
Ecoli	10(11)	10(11)	9(12)	10(11)	10(13)	10(12)	10(12)	10(11)	10(11)	10(12)
GI	2(12)	2(27)	2(28)	2(27)	2(30)	2(30)	2(31)	2(30)	2(30)	2(31)
Haberman	4(8)	4(7)	3(11)	3(16)	3(12)	3(10)	3(15)	3(17)	3(22)	3(17)
Heart Statlog	10(13)	10(8)	10(10)	9(9)	8(8)	10(11)	10(13)	9(11)10(11)	10(10)	9(17)
Ionosphere	10(8)	6(8)7(8)	8(8)	9(9)10(9)	10(9)	8(9)9(9)	7(9)	8(11)	7(9)8(9)	8(7)10(7)
Iris	2(29)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Lymph	10(27)	10(20)	10(19)	10(25)	10(26)	10(20)	10(23)	10(24)	10(18)	10(20)
MB	10(29)	10(31)	10(26)	10(28)	10(29)	10(29)	10(26)	10(29)	10(28)	10(28)
MF	10(23)	10(18)	10(24)	10(28)	10(26)	10(29)	10(21)	10(27)	10(24)	10(23)
Parkinsons	2(15)	2(19)	2(22)	2(24)	2(24)	2(21)	2(28)	2(28)	2(30)	2(31)
Pima	2(27)	2(28)	2(29)	2(26)	2(28)	2(29)	2(30)	2(31)	2(31)	2(30)
Planning Relax	10(20)	10(22)	10(21)	10(23)	10(20)	10(19)	10(22)	10(20)	10(22)	10(24)
QSAR-B	10(13)	9(11)	10(15)	10(12)	10(11)	10(12)	9(12)	10(15)	10(18)	10(11)
Robot Failure	10(30)	10(19)	10(16)	10(13)	10(13)	10(10)	10(14)	9(8)10(8)	10(11)	8(11)
Seeds	7(8)	5(6)	10(8)	9(7)10(7)	9(8)	10(8)	10(9)	9(9)	8(12)	10(13)
Semeion	10(27)	10(26)	10(28)	10(27)	10(24)	10(28)	10(31)	10(26)	10(28)	10(26)
Sonar	10(20)	10(17)	10(22)	10(18)	10(20)	10(23)	10(20)	10(20)	10(21)	10(21)
SPF	2(28)	2(30)	2(29)	2(31)	2(31)	2(30)	2(31)	2(31)	2(31)	2(31)
TS	2(26)	2(30)	2(31)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
UKM	10(18)	10(15)	10(19)	10(18)	10(18)	10(20)	10(16)	10(22)	10(15)	10(17)
Vehicle	2(30)	2(31)	2(29)	2(31)	2(31)	2(28)	2(30)	2(31)	2(31)	2(31)
Voice	2(28)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Wine	4(7)10(7)	3(15)	3(15)	3(19)	3(16)	3(19)	4(15)	4(14)	3(15)	4(30)

Table 7: ckMeans-Sugeno: Number of clusters selected for multiple subsets of data

Datasets	**Percentages of Data**									
	10%	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Balance	10(17)	10(15)	10(19)	10(18)	10(15)	10(24)	10(22)	10(22)	10(26)	10(25)
GI	9(11)10(11)	10(14)	10(18)	10(14)	10(15)	9(15)	9(19)	9(17)	9(18)	9(17)
Bupa	9(8)10(8)	2(12)	2(14)	2(15)	2(15)	2(12)	2(14)	2(13)	2(17)	4(24)
CMSC	9(7)	6(9)	9(9)	5(14)	5(9)	5(15)	5(20)	5(23)	5(25)	5(27)
Cnae-9	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Column 3C	10(12)	10(7)	4(15)	4(15)	4(13)	4(16)	4(16)	4(16)	3(19)	3(29)
Contraceptive	10(11)	10(17)	10(16)	10(22)	10(17)	10(19)	10(20)	10(17)	10(21)	10(20)
Ecoli	10(16)	10(12)	10(12)	6(9)	7(7)8(7)10(7)	7(10)	8(9)	7(8)9(8)	9(10)	7(10)
GI	10(9)	10(8)	2(7)4(7)	4(10)	2(8)	4(10)	4(13)	4(16)	4(13)	4(24)
Haberman	10(13)	10(9)	10(11)	9(11)	10(9)	10(11)	9(11)	10(13)	10(18)	10(11)
Heart Statlog	10(15)	10(12)	8(8)	8(7)	8(7)9(7)	8(7)	7(7)8(7)	8(9)	8(9)	5(10)
Ionosphere	10(19)	10(14)	8(8)	10(11)	8(8)9(8)	8(11)	7(10)	8(10)	9(9)	8(9)
Iris	2(20)	2(29)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Lymph	10(29)	10(19)	10(27)	10(14)	10(18)	10(22)	10(26)	10(23)	10(19)	10(22)
MB	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
MF	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Parkinsons	3(7)	3(11)	4(11)	4(11)	4(15)	4(12)	4(16)	4(19)	4(26)	4(31)
Pima	6(9)	4(7)	5(11)	5(11)	5(12)	5(14)	5(17)	5(19)	5(21)	5(23)
Planning Relax	10(26)	10(23)	10(24)	10(25)	10(25)	10(28)	10(27)	10(29)	10(29)	10(29)
QSAR-B	9(6)	5(8)	7(9)	6(9)	6(9)	5(15)	7(9)	5(9)6(9)	6(11)	5(16)
Robot Failure	10(31)	10(24)	10(23)	10(18)	10(18)	10(14)	10(16)	10(13)	9(13)	10(16)
Seeds	10(19)	10(12)	10(9)	10(9)	6(6)7(6)10(6)	10(11)	8(8)	6(9)	8(7)	10(10)
Semeion	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Sonar	10(24)	10(24)	10(21)	10(21)	10(23)	10(16)	10(23)	10(28)	10(20)	10(25)
SPF	4(17)	4(22)	4(15)	4(19)	5(17)	5(15)	5(18)	5(28)	5(29)	5(31)
TS	2(27)	2(30)	2(31)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
UKM	10(20)	10(15)	10(19)	10(23)	10(19)	10(17)	10(22)	10(24)	10(27)	10(23)
Vehicle	2(12)	2(17)	4(14)	4(16)	4(20)	4(22)	4(22)	4(21)	4(26)	4(28)
Voice	4(10)	6(15)	6(16)	6(22)	6(18)	6(21)	6(23)	6(22)	6(28)	6(31)
Wine	10(19)	4(8)	5(7)10(7)	5(10)	5(11)	6(10)	5(13)	5(16)	5(12)	4(30)

Table 8: FCMσ-MPC: Number of clusters selected for multiple subsets of data

Datasets	**Percentages of Data**									
	10%	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Balance	3(15)	3(15)	2(19)	2(18)	2(24)	2(27)	2(27)	2(28)	2(31)	2(30)
GI	2(28)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Bupa	2(19)	2(23)	2(13)	2(15)	2(14)	2(15)	2(11)	2(8)	3(7)4(7)	5(8)
CMSC	3(7)5(7)	4(10)	3(15)	3(19)	3(19)	3(25)	3(27)	3(31)	3(31)	3(31)
Cnae-9	7(6)	10(8)	8(8)10(8)	10(7)	8(8)	9(10)	9(10)	10(8)	8(8)	8(8)
Column 3C	2(30)	2(31)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Contraceptive	2(24)	2(21)	2(22)	2(24)	2(25)	2(28)	2(24)	2(20)	2(26)	2(25)
Ecoli	2(12)	10(8)	2(12)	2(14)	2(16)	3(15)	2(18)	2(18)	3(16)	2(17)
GI	2(23)	2(30)	2(30)	2(29)	2(30)	2(31)	2(30)	2(30)	2(31)	2(31)
Haberman	2(10)	2(23)	2(27)	2(24)	2(30)	2(29)	2(31)	2(31)	2(31)	2(31)
Heart Statlog	2(14)	2(23)	2(27)	2(29)	2(31)	2(29)	2(30)	2(31)	2(31)	2(31)
Ionosphere	2(13)	2(22)	2(24)	2(25)	2(24)	2(25)	2(27)	2(29)	2(28)	2(28)
Iris	2(16)	2(20)	2(20)	2(18)	2(18)	2(22)	2(19)	2(20)	2(22)	2(31)
Lymph	3(14)	2(13)	3(14)	2(13)3(13)	2(15)	3(18)	2(14)	2(14)	2(15)	2(17)
MB	5(18)	7(12)	7(13)	7(13)	8(11)	9(10)	8(10)	7(10)	7(12)	8(8)
MF	3(7)	2(9)	2(8)	2(10)	2(12)	2(14)	2(13)	2(15)	2(13)	2(14)
Parkinsons	2(26)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Pima	2(31)	2(31)	2(31)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Planning Relax	3(9)	3(12)	3(10)	2(11)	2(11)	2(11)	2(9)	2(9)	2(8)3(8)	3(11)
QSAR-B	2(29)	2(31)	2(31)	2(30)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)
Robot Failure	10(31)	2(16)	2(15)	2(17)	2(24)	2(25)	2(17)	2(24)	2(22)	2(22)
Seeds	2(16)	2(26)	2(27)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(30)
Semeion	10(24)	9(16)	10(19)	10(18)	10(17)	10(15)	10(15)	10(17)	9(14)	9(12)
Sonar	3(16)	2(13)	3(11)	3(13)	2(12)3(12)	2(14)	2(9)	2(11)	3(10)	2(8)
SPF	2(26)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
TS	2(12)	2(14)	10(15)	10(16)	10(24)	10(21)	10(22)	10(28)	10(29)	10(31)
UKM	2(13)3(13)	2(14)	3(13)	2(14)	2(14)	2(13)	2(12)3(12)	2(13)	2(13)	3(10)
Vehicle	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Voice	2(27)	2(30)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Wine	2(13)	2(23)	2(20)	2(24)	3(16)	2(22)	2(20)	2(18)	2(21)	2(31)

Table 9: FCMσ-XB: Number of clusters selected for multiple subsets of data

Datasets	**Percentages of Data**									
	10%	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Balance	3(8)	10(7)	10(8)	10(13)	9(11)	10(15)	10(17)	10(15)	10(11)	10(13)
GI	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Bupa	2(20)	2(26)	2(19)	2(24)	2(22)	2(24)	2(21)	2(20)	2(21)	2(25)
CMSC	2(23)	2(24)	2(29)	2(27)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)
Cnae-9	2(8)	9(6)10(6)	9(7)	10(6)	2(8)	9(6)10(6)	10(8)	10(7)	8(7)	9(9)
Column 3C	2(31)	2(31)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Contraceptive	2(18)	2(19)	2(19)	2(16)	2(21)	2(19)	2(20)	2(19)	2(24)	2(23)
Ecoli	2(16)	2(16)	2(16)	2(15)	2(23)	2(21)	2(24)	2(22)	2(24)	2(21)
GI	2(26)	2(31)	2(31)	2(29)	2(30)	2(31)	2(30)	2(30)	2(31)	2(31)
Haberman	2(13)	2(15)	2(18)	2(17)	2(21)	2(23)	2(20)	2(19)	2(16)	2(28)
Heart Statlog	2(17)	2(21)	2(23)	2(25)	2(28)	2(25)	2(30)	2(28)	2(27)	2(28)
Ionosphere	2(22)	2(21)	2(17)	2(23)	2(21)	2(22)	2(25)	2(26)	2(27)	2(22)
Iris	2(28)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Lymph	2(8)	2(9)	2(11)	2(10)	2(9)	2(10)3(10)	2(12)	2(12)	2(13)	2(16)
MB	5(19)	6(12)7(12)	7(13)	8(16)	8(12)	9(11)	8(12)9(12)	7(10)8(10)	7(11)	8(12)
MF	2(7)	2(15)	2(8)	2(12)	2(13)	2(13)	2(11)	2(11)	2(13)	2(18)
Parkinsons	2(27)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Pima	2(31)	2(31)	2(31)	2(30)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Planning Relax	3(8)	4(9)	9(7)	2(6)9(6)	10(12)	10(10)	9(8)	7(8)	9(10)	10(12)
QSAR-B	2(29)	2(31)	2(30)	2(30)	2(28)	2(30)	2(31)	2(31)	2(31)	2(31)
Robot Failure	10(31)	2(19)	2(13)	2(16)	2(24)	2(19)	2(20)	2(20)	2(22)	2(25)
Seeds	2(18)	2(31)	2(29)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Semeion	10(14)	10(16)	9(14)	10(17)	10(19)	10(17)	10(18)	9(14)	10(13)	10(16)
Sonar	2(12)	2(11)	2(8)	2(11)	3(8)	2(9)	3(9)	3(8)	2(11)	2(11)
SPF	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
TS	2(24)	2(24)	2(27)	2(25)	2(28)	2(27)	2(29)	2(28)	2(25)	2(30)
UKM	2(12)	2(8)4(8)	2(12)	2(9)	2(7)3(7)	2(14)	2(9)	2(11)	2(10)	3(10)
Vehicle	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)	2(31)
Voice	2(23)	2(28)	2(27)	2(30)	2(29)	2(31)	2(31)	2(31)	2(31)	2(31)
Wine	2(15)	2(14)	3(17)	3(15)	3(25)	3(28)	3(30)	3(30)	3(29)	3(31)

Table 10: FCMσ-Sugeno: Number of clusters selected for multiple subsets of data

Datasets	**Percentages of Data**									
	10%	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Balance	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
GI	10(25)	10(24)	10(26)	10(30)	10(31)	10(29)	10(29)	10(28)	10(29)	10(28)
Bupa	10(28)	10(30)	10(30)	10(31)	10(30)	10(31)	10(31)	10(31)	10(31)	10(31)
CMSC	6(8)	6(7)	5(8)7(8)	5(11)	5(15)	5(14)	5(23)	5(17)	5(26)	5(31)
Cnae-9	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Column 3C	10(16)	10(25)	10(25)	10(26)	10(27)	10(26)	10(29)	10(30)	10(26)	10(30)
Contraceptive	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Ecoli	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
GI	10(22)	10(20)	10(24)	10(28)	10(28)	10(31)	10(30)	10(30)	10(31)	10(31)
Haberman	10(21)	10(18)	10(20)	10(23)	10(24)	10(26)	10(28)	10(31)	10(31)	10(30)
Heart Statlog	10(27)	10(30)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Ionosphere	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Iris	2(12)	3(15)	3(24)	3(23)	3(28)	3(27)	3(29)	3(31)	3(31)	3(31)
Lymph	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
MB	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
MF	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Parkinsons	10(9)	2(9)	3(8)	2(8)	3(8)	3(8)	5(8)	5(7)10(7)	5(17)	5(25)
Pima	2(15)	2(17)	2(17)	2(13)	2(15)	2(25)	2(22)	2(25)	2(28)	2(31)
Planning Relax	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
QSAR-B	10(29)	10(31)	10(30)	10(30)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Robot Failure	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Seeds	10(16)	10(13)	10(14)	10(22)	10(25)	10(22)	10(24)	10(26)	10(28)	10(28)
Semeion	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Sonar	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
SPF	6(11)	6(18)	5(16)	5(15)6(15)	6(19)	6(17)	6(22)	6(23)	6(24)	6(31)
TS	10(22)	10(31)	10(29)	10(30)	10(31)	10(31)	10(30)	10(31)	10(31)	10(31)
UKM	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)	10(31)
Vehicle	3(17)	3(24)	3(30)	3(30)	3(30)	3(31)	3(31)	3(31)	3(31)	3(31)
Voice	3(23)	3(23)	3(25)	3(25)	3(23)	3(30)	3(31)	3(30)	3(31)	3(31)
Wine	10(10)	4(10)	4(16)	4(19)	4(23)	4(24)	4(28)	4(30)	4(31)	4(31)

From Tables 2, 3 and 4, we can observe that the FCM algorithm showed similar behavior with 15, 12 and 15 datasets, respectively. In other words, the same number of clusters was chosen in all subsets and in the total set, when using the MPC, XB and Sugeno index for evaluation. The ckMeans algorithm showed similar behavior with 19, 22 and 15 datasets, when using the MPC, XB and Sugeno index, respectively, as shown in Tables 5 to 7. The FCMσ algorithm showed similar behavior with 15, 20 and 25 datasets, considering the MPC, XB and Sugeno index, respectively, according to Tables 8 to 10. Considering that FCM, ckMeans and FCMσ are non-deterministic algorithms, in some cases there was a difference in the number of times the number of clusters was chosen.

From Tables 2 to 10, we can observe that, in a general perspective, the analyzed algorithms showed similar behavior, selecting the same number of clusters for almost all the data subsets as well as for the original dataset. It is important to emphasize that there was a discrepancy in some scenarios when selecting the number of clusters. Therefore, it is necessary to apply statistical tests in order to assess the performance delivered by the analyzed scenarios, from a statistical point of view.

4.1 Statistical Test

In the statistical analysis, the Kruskal-Wallis test as well as the Mann-Whitney test were used. The Kruskal-Wallis test was used to compare the behavior of the algorithms in each data subset and in the original dataset. Therefore, it is applied directly to the classification result, i.e., over a vector of 31 positions, where each position refers to the number of clusters selected in the corresponding execution. This test serves to evaluate the hypothesis that the different percentages of data have the same distribution.

The results of the Kruskal-Wallis test for the FCM, ckMeans and FCM σ algorithms are presented in Table 11. The values presented in each cell correspond to the p-value of this test for each dataset, algorithm and validation index.

Table 11: p-value result from the Kruskal Wallis test

Datasets	**FCM**			**ckMeans**			**FCMσ**		
	MPC	**XB**	**Sugeno**	**MPC**	**XB**	**Sugeno**	**MPC**	**XB**	**Sugeno**
Balance	0.0094	0.0191	0.0255	0.0314	<0.0001	0.0027	<0.0001	<0.0001	0.5000
GI	0.5000	0.4687	0.9397	0.5000	0.5543	0.4453	0.0013	0.4373	0.0526
Bupa	<0.0001	<0.0001	<0.0001	<0.0001	0.5248	<0.0001	0.0001	0.3967	0.1144
CMSC	<0.0001	<0.0001	0.0996	<0.0001	0.0122	<0.0001	0.0002	<0.0001	<0.0001
Cnae-9	0.3136	0.3520	0.5000	0.2736	0.7996	0.5000	0.0867	0.1878	0.5000
Column 3C	0.0115	<0.0001	<0.0001	0.5315	0.0014	<0.0001	0.5315	0.4373	<0.0001
Contraceptive	0.0027	0.0512	<0.0001	0.1313	0.7323	0.0814	0.3865	0.6837	0.5000
Ecoli	<0.0001	0.2498	<0.0001	0.0270	0.9997	<0.0001	0.0002	0.0195	0.5000
GI	<0.0001	0.0005	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0078	<0.0001
Haberman	<0.0001	0.0436	0.0002	<0.0001	0.1271	0.1774	<0.0001	0.0002	<0.0001
Heart Statlog	<0.0001	0.5155	0.0025	<0.0001	0.0160	<0.0001	<0.0001	0.0002	0.0006
Ionosphere	<0.0001	<0.0001	0.1281	<0.0001	0.2875	0.0002	<0.0001	0.2473	0.5000
Iris	0.0345	0.0013	<0.0001	0.5000	0.0345	<0.0001	0.0035	0.0013	0.0303
Lymph	<0.0001	<0.0001	0.0318	<0.0001	0.0706	0.0002	0.1517	0.5758	0.5000
MB	<0.0001	0.3280	0.5000	<0.0001	0.5213	0.5000	<0.0001	<0.0001	0.5000
MF	0.0519	0.0010	0.5000	0.0001	0.0214	0.5000	<0.0001	0.0180	0.5000
Parkinsons	0.0004	<0.0001	<0.0001	<0.0001	<0.0001	0.1569	<0.0001	<0.0001	0.2060
Pima	0.5000	0.0116	0.0024	0.5000	0.1914	0.0594	0.4373	0.4373	<0.0001
Planning Relax	<0.0001	<0.0001	0.5315	<0.0001	0.9457	0.2321	0.6555	<0.0001	0.5000
QSAR-B	0.5000	0.5000	0.0463	0.4373	0.4376	0.0114	0.2684	0.1981	0.2684
Robot Failure	<0.0001	<0.0001	0.0004	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.5000
Seeds	<0.0001	<0.0001	0.0006	<0.0001	0.0073	<0.0001	<0.0001	<0.0001	<0.0001
Semeion	0.0002	0.4946	0.5000	0.0004	0.4341	0.5000	0.0005	0.6987	0.5000
Sonar	<0.0001	<0.0001	0.0088	<0.0001	0.9374	0.0964	0.1381	0.4885	0.5000
SPF	<0.0001	<0.0001	<0.0001	<0.0001	0.0974	<0.0001	<0.0001	0.5000	<0.0001
TS	<0.0001	<0.0001	<0.0001	0.4373	0.0002	0.0034	<0.0001	0.2740	<0.0001
UKM	<0.0001	<0.0001	0.0974	<0.0001	0.4562	0.0407	0.6245	0.5412	0.5000
Vehicle	0.5000	0.5000	<0.0001	0.5000	0.1034	0.1338	0.4373	0.5000	<0.0001
Voice	<0.0001	0.5000	<0.0001	<0.0001	0.0013	0.0085	0.0037	<0.0001	<0.0001
Wine	0.0007	0.2636	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0045	<0.0001

Based on the values presented in Table 11, it can be observed that there is a statistically significant difference in all validation algorithms and indices (p-value ≤ 0.05). Therefore, it is necessary to identify which subsets showed a statistically significant difference in relation to the total set. In this context, Mann-Whitney test was applied for all algorithms and validation indices. Mann-Whitney test is a non-parametric method well known in the literature, which compares two paired samples, and can identify pairs that are statistically different.

Table 12 presents the results of Mann-Whitney test when comparing the behavior obtained by the algorithms in each subset of data with the total set. In this analysis, we took into account the number of datasets that did not present a statistically significant difference by percentage of data, for each validation index and for each clustering algorithm. For example, using the FCM algorithm, MPC index and 40% of the data from the total set, we have 21 datasets (out of 30), which did not differ from a statistical point of view. Therefore, values outside parentheses indicate the number of datasets that did not show a statistically significant difference, while values inside parentheses indicate otherwise.

Table 12: Result of the Mann-Whitney test when comparing each subset of data with the total set

Algorithms	Indices	10%	20%	30%	40%	50%	60%	70%	80%	90%
FCM	**MPC**	7(23)	14(16)	19(11)	21(9)	24(6)	23(7)	25(5)	25(5)	27(3)
	XB	12(18)	17(13)	18(12)	21(9)	25(5)	22(8)	24(6)	23(7)	27(3)
	Sugeno	13(17)	14(16)	16(14)	17(13)	15(15)	19(11)	19(11)	20(10)	22(8)
ckMeans	**MPC**	11(19)	15(15)	19(11)	20(10)	21(9)	23(7)	22(8)	24(6)	25(5)
	XB	18(12)	19(11)	23(7)	22(8)	25(5)	26(4)	26(4)	27(3)	28(2)
	Sugeno	10(20)	17(13)	18(12)	20(10)	20(10)	25(5)	24(6)	26(4)	23(7)
FCMσ	**MPC**	6(24)	15(15)	19(11)	19(11)	22(8)	23(7)	25(5)	26(4)	27(3)
	XB	11(19)	19(11)	22(8)	21(9)	26(4)	26(4)	26(4)	27(3)	28(2)
	Sugeno	17(13)	20(10)	20(10)	20(10)	22(8)	22(8)	26(4)	27(3)	26(4)

Table 13 presents the percentage of datasets that did not show any difference from the statistical point of view, when comparing each data subset (column) with the original dataset. The percentage was calculated based on the value obtained in each index for each algorithm. For example, in order to calculate the percentage of datasets that did not show statistical difference taking into account $p = 10\%$ and the FCM algorithm, we took into consideration the values presented in each index and algorithm ($7 + 12 + 13 = 32$) and the percentage of the value is calculated, taking into account all 30 datasets used in the experiments, for each index and algorithm (3 indices x 30 datasets = 90). Therefore, the percentage of datasets that did not show statistical difference for the FCM algorithm when 10% of the data was used is equal to 35.56% ((32x100)/90 = 35.56).

Table 13: Percentage of Datasets that did not show statistical difference

Algorithms	Percentage of Data								
	10%	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**
FCM	35.56%	50.00%	58.89%	65.56%	71.11%	71.11%	75.56%	75.56%	84.44%
ckMeans	43.33%	56.67%	66.67%	68.89%	73.33%	82.22%	80.00%	85.56%	84.44%
FCMσ	37.78%	60.00%	67.78%	66.67%	77.78%	78.89%	85.56%	88.89%	90.00%

In this analysis, let consider percentages equal or higher than 70% was considered as a strong percentage (more than 2/3 of the subsets that did not present a statistically significant difference). The use of the strong percentage concept aims to identify the smallest data subset that delivers similar perfomance than the orginal dataset (reaches the strong percentage). Therefore, from Table 13, we can observe that the strong percentage was reached with $p \geq 50\%$, for all the analyzed algorithms (shadded values).

Based on this analysis, it is possible to infer that the number of clusters in a dataset can be done using a data subset higher than 50% ($p \geq 50\%$). It is important to highlight that this reduction in the number of instances in a dataset allows to optimize the computational time and processing cost in the data clustering process, especially when using large datasets with no previous knowledge about the data.

5 Final Remarks

Several approaches in the literature investigate the problem of defining the number of clusters in a dataset. In general, these approaches use the original dataset as a way to determine the number of clusters. In this work, we investigated the use of data subsets in the definition of the number of clusters. It is a smaller sample, but still able to infer the number of clusters of a dataset. In this investigation, three clustering algorithms (FCM, ckMeans and FCMσ) were applied and assessed using three validation indices. In order to assess the performance of this proposal, we performed an empirical analysis with 30 datasets. Each dataset was partitioned into 9 data subsets, starting with 10% ($p = 10\%$) and increasing with 10% intervals untils it reaches the original dataset size (p=10, 20, ...,90%).

Through the empirical analysis, we can conclude that the results obtained in this article are promising, both from an experimental point of view and from a statistical point of view. These results show that the use of a smaller percentage of a dataset can be used to infer the number of clusters with an efficient performance. More specifically, data subsets higher than 50% ($p \geq 50\%$) present results similar to the original dataset when defining the best number of clusters of a dataset.

In future work, we can use a larger number of datasets and characterize them categorically, in an attempt to find the data percentage to infer the number of clusters, according to the characteristics of each dataset. In addition, we can use other clustering algorithms, such as k-Means and initialization of initial centers as in[32], and perform a similar study for interval fuzzy clustering algorithms [33] or use this methodology applied to the context of clustering ensembles optimization [34].

Conflict of Interest: The authors declare no conflict of interest.

Funding: ”This work has been partially supported by Brazilian Coordination of Superior Level Staff Improvement CAPES and Brazilian National Council of Technological and Scientific Development CNPq (Proc. 311429/2020-3).”

References

[1] Carvalho A, Faceli K, Lorena A, Gama J. *Inteligência Artificial–uma Abordagem de Aprendizado de Máquina.* Rio de Janeiro: LTC; 2011.

[2] Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL. *Análise Multivariada de Dados.* Porto Alegre: Bookman; 2009.

[3] Bong CW, Rajeswari M. Multi-objective nature-inspired clustering and classification techniques for image segmentation. *Applied soft computing.* 2011; 11(4): 3271-3282. DOI: https://doi.org/10.1016/j.asoc.2011.01.014

[4] Sun H, Wang S, Jiang Q. FCM-based model selection algorithms for determining the number of clusters. *Pattern recognition.* 2004; 37(10): 2027-2037. DOI: https://doi.org/10.1016/j.patcog.2004.03.012

[5] Sentelle C, Hong SL, Anagnostopoulos GC, Georgiopoulos M. A Fuzzy GAP statistic for Fuzzy c-means. In: International Conference Artificial Intelligence and Soft Computing. 2007.

[6] Kodinariya TM, Makwana PR. Review on determining number of cluster in K-means clustering. *International Journal.* 2013; 1(6): 90-95. DOI:10.18576/amis/100428

[7] Xu S, Hu L, Yang X, Liu X. A cluster number adaptive Fuzzy c-means algorithm for image segmentation. *International Journal of Signal Processing, Image Processing and Pattern Recognition.* 2013; 6(5): 191-204. DOI: http://dx.doi.org/10.14257/ijsip.2013.6.5.17

[8] Ren M, Liu P, Wang Z, Yi J. A self-adaptive Fuzzy c-means algorithm for determining the optimal number of clusters. *Computational intelligence and neuroscience.* 2016; 2016(3): 1-12. DOI: https://doi.org/10.1155/2016/2647389

[9] Ünlü R, Xanthopoulos P. Estimating the number of clusters in a dataset via consensus clustering. *Expert Systems with Applications.* 2019; 125: 33-39. DOI: https://doi.org/10.1016/j.eswa.2019.01.074

[10] Jain AK, Murty MN, Flynn PJ. Data clustering: A review.*ACM computing surveys (CSUR).* 1999; 31(3): 264-323. DOI: https://doi.org/10.1145/331499.331504

[11] Wang W, Zhang Y. On Fuzzy cluster validity indices. *Fuzzy sets and systems.* 2007; 158(19): 2095-2117. DOI: https://doi.org/10.1016/j.fss.2007.03.004

[12] Xing R, Li C. Fuzzy c-means algorithm automatically determining optimal number of clusters. *Computers, Materials and Continua.* 2019; 60(2): 767-780. DOI: https://doi.org/10.32604/cmc.2019.04500

[13] Yejun X. Optimization of the clusters number of an improved fuzzy C-means clustering algorithm. In: 2015 10th International Conference on Computer Science & Education (ICCSE). 2015. p.931-935. DOI: 10.1109/ICCSE.2015.7250383

[14] Dinh DT, Fujinami T, Huynh VN. Estimating the optimal number of clusters in categorical data clustering by silhouette coefficient. In: Knowledge and Systems Sciences. 2019. p.1-17. DOI: https://doi.org/10.1007/978-981-15-1209-4_1

[15] Liao HY, Ng MK. Categorical data clustering with automatic selection of cluster number. *Fuzzy Information and Engineering.* 2009; 1(1): 5-25. DOI: https://doi.org/10.1007/s12543-009-0001-5

[16] Havens TC, Bezdek JC, Palaniswami M. Cluster validity for kernel Fuzzy clustering. In: 2012 IEEE International Conference on Fuzzy Systems. 2012. p.1-8. DOI: https://doi.org/10.1109/FUZZ-IEEE.2012.6250820

[17] Markelle K, Rachel L, Kolby N. The UCI Machine Learning Repository, https://archive.ics.uci.edu

[18] Smith JW, Everhart JE, Dickson WC, Knowler WC, Johannes RS. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In: Proceedings of the annual symposium on computer application in medical care. 1988. p.261.

[19] Bezdek JC. *Pattern Recognition with Fuzzy Objective Function Algorithms.* Norwell, MA, USA: Kluwer Academic Publishers; 1981.

[20] Dunn JC. A Fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics.* 1973; 3(3): 32-57. DOI: https://doi.org/10.1080/01969727308546046

[21] Bedregal BR. A comparative study between fuzzy c-means and ckmeans algorithms. In: 2010 Annual Meeting of the North American Fuzzy Information Processing Society. 2010. p.1-6. DOI: https://www.doi.org/10.1109/NAFIPS.2010.5548194

[22] Tsai DM, Lin CC. Fuzzy C-means based clustering for linearly and nonlinearly separable data. *Pattern recognition.* 2011; 44(8): 1750-1760. DOI: https://doi.org/10.1016/j.patcog.2011.02.009

[23] Bezdek JC, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm. *Computers & geosciences.* 1984; 10(2-3): 191-203. DOI: https://doi.org/10.1016/0098-3004(84)90020-7

[24] Vargas RR, Bedregal BR, Palmeira ES. A comparison between k-means, fcm and ckmeans algorithms. In: 2011 Workshop-School on Theoretical Computer Science. 2011. p.32-38. DOI: https://doi.org/10.1109/WEIT.2011.28

[25] Vargas RR, Freddo R, Galafassi C, Gass SL, Russini A, Bedregal B. Identifying pixels classified uncertainties ckmeansimage algorithm. In: Information Processing and Management of Uncertainty in Knowledge-Based Systems. 2018. p.429-440. DOI:10.1007/978-3-319-91479-4_36

[26] Jain AK, Dubes RC. *Algorithms for Clustering Data.* Prentice-Hall; 1988.

[27] Halkidi M, Batistakis Y, Vazirgiannis M. On clustering validation techniques. *Journal of intelligent information systems.* 2001; 17(2-3): 107-145. DOI: https://doi.org/10.1023/A:1012801612483

[28] Peres SM, Rocha T, Bíscaro HH, Madeo RC, Boscarioli C. Tutorial sobre Fuzzy-c-means e Fuzzy learning vector Quantization: Abordagens híbridas para tarefas de agrupamento e classificação. *Revista de Informática Teórica e Aplicada.* 2012; 19(1): 120-163. DOI: https://doi.org/10.22456/2175-2745.13764

[29] Dave RN. Validating Fuzzy partitions obtained through c-shells clustering. *Pattern recognition letters.* 1996; 17(6): 613-623. DOI: https://doi.org/10.1016/0167-8655(96)00026-8

[30] Fukuyama Y, Sugeno M. A new method of choosing the number of clusters for the Fuzzy c-mean method. *The 5th Fuzzy Syst Symp.* 1989; 247: 247-250.

[31] Bezdek JC, Pal NR. Some new indexes of cluster validity. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics).* 1998; 28(3): 301-315. DOI: https://doi.org/10.1109/3477.678624

[32] da Silva LR, Arnaldo HA, da Silva H, Moura R, Bedregal B, de P Canuto AM. Two deterministic selection methods for the initial centers in Fuzzy c-means based algorithms. *Intelligent Data Analysis.* 2020; 24(4): 779-798. DOI: https://doi.org/10.3233/IDA-194588

[33] Silva L, Moura R, Canuto AM, Santiago RH, Bedregal B. An interval-based framework for uzzy clustering applications. *IEEE Transactions on Fuzzy Systems.* 2015; 23(6): 2174-2187. DOI: https://doi.org/10.1109/TFUZZ.2015.2407901

[34] Silva HM, Canuto AM, Medeiros IG, Xavier-Júnior JC. Cluster ensembles optimization using coral reefs optimization algorithm. In: Artificial Neural Networks and Machine Learning–ICANN 2016: 25th International Conference on Artificial Neural Networks. 2016; p.275-282. DOI: https://doi.org/10.1007/978-3-319-44781-0_33

Huliane Medeiros da Silva
Departamento de Engenharias e Tecnologias
Universidade Federal Rural do Semi-Árido
Pau dos Ferros, Brasil

E-mail: huliane@ufersa.edu.br

Benjamín René Callejas Bedregal
Departamento de Informática e Matemática Aplicada
Universidade Federal do Rio Grande do Norte
Natal, Brasil
E-mail: bedregal@dimap.ufrn.br

Anne Magály de Paula Canuto
Departamento de Informática e Matemática Aplicada
Universidade Federal do Rio Grande do Norte
Natal, Brasil
E-mail: anne.canuto@ufrn.br

Thiago Vinícius Vieira Batista
Departamento de Informática e Matemática Aplicada
Universidade Federal do Rio Grande do Norte
Natal, Brasil
E-mail: thiagovvb@gmail.com

Ronildo Pinheiro de Araújo Moura
Departamento de Informática e Matemática Aplicada
Universidade Federal do Rio Grande do Norte
Natal, Brasil
E-mail: ronildoamoura@gmail.com