JOURNAL OF
PRODUCTION AND AUTOMATION

v.2, n.1, p.41-60, 2019.

Aerotaxonomy

Eden Denis Ferreira da Silva Lopes Santos [1], Carlos Alberto Rocha Pimentel[1] and Aguinaldo Prandini Ricieri [2]

[1]Federal University of ABC (UFABC)

[2] Technological Institute of Aeronautics (ITA)

ABSTRACT

This work aims to classify military aircraft into possible groups, each of which must contain aircraft with some categorical similarity according to an Euclidean Distance Function, evolutionary allegory between the various types of aircrafts. This is a Numerical Taxonomy or Computational Analysis of Clustering/Data Mining, for this we choose *priori*, the total number of K groups that will dendrogram the aircraft of low, medium, high efficiency in one feature; which could serve as an Informational Input for new aircraft pre-projects.

Keywords: Aircraft Taxonomy, Aircraft Classification for Pre-design, Aircraft Big Data, Dendrogram, k-means, Python, Artificial Intelligence.

INTRODUCTION

The problem of partitioning has great expression in the area of Biology as well, which is why scientists have created a specific branch called taxonomy for this area of knowledge, that of organisms and plants. However, all of this has come from the creation of societies when Aristotle, in about $300a.C$, initially and generally separated the living beings from the plants and later on, for the case of the first group, for example, separated them by blood color, between red and not. In the XVIII century, Carl von Linné $(1707 - 1778)$ improved the taxonomic process (used to date) that better organized living beings, since more species were known. Currently partitioned into five kingdoms: Monera, Protista, Fungi, Animalia and Plantae; Linné divided living beings by taking into account seven criteria (kingdom, phylum, class, order, family, gender and species), the first criterion is more generalist and, as the criterion advances, more specific analysis is made and soon, a better classification is gotten. Since the objective of this work is to classify military aircraft, the neologism Aerotaxonomy becomes quite adequate, once, the word aero is related to aerospace engineering and, therefore, taxonomy means classification [1], given that the term Biotaxonomy would be convenient in biology. Recently, a domain has been added before the kingdom. Providing that the word aero is comprehensive, it is convenient to use, in this case, Aerotaxonomy of aircraft as it, could also be of engines.

Big Data, in particular its two techniques or methods for processing data, k -means and Dendrogramming, can serve as a model for modern knowledge characterized by its limited temporal validity. And mathematics such as that of the XX century, which runs counter to this particularity of the modern world, the XXI century, Bauman's liquid society [2], is useless for the purposes of any institution of current and likely future scientific-technological knowledge.

For data scientists, Big Data and its techniques is what should be studied and applied in modern engineering, particulary in aeronautics; not because it offers a new field for the knowledge investigation, as it has happened with classical mathematics so far, but rather because its predictive results encourage the elevation of Big Data's status as the main discipline that can govern engineering, which must necessarily be concerned with aspects never before thought of by engineers as popular preferences, sales plans, marketing options, success percentages and product form subjectivities; which is also required by aeronautical products.

Whereas the traditional engineer seeks, at all costs, a way of imposing theories on the phenomena of the world that can fully explain them; the data engineer analyzes them as they present themselves to us, with some degree of ambiguity. This new, pragmatic approach, which tolerates possible errors or deviations from prediction, from the semiotic-abdutional point of view, allegorizes two predictions movements, of the same direction and inverse senses: of the models, with their syntactic symbols to the data and with their computational mining to the models. Hence the essence of the dissonance between classical engineering, first sense, and data engineering, second sense, is in the complete theoretical explanations and in the limited computational models.

For a data engineer, a professional of an immediate nature, who works in an increasingly liquid society, the search for a perfect, official and universally comprehensive model forbids him from any possibility of freely predicting possible minimally assertive results. Thus, the essence of the conflict between these two exact professionals displaced in time, slips in the diversity of knowledge produced in many sectors of this or that Engineering, which require quick and objective answers, but not extremely precise ones.

So, knowing that Big Data, as mathematics that has influenced as never before engineering, allows countless answers, what remains for the ones who call themselves engineers or data scientist to know what the questions are.

Therefore, this modern mathematics, adjusted to the present society and that has defined new solution standards added to the taxonomy, the main objective of this work is to describe a process to partition a d -dimensional population of aircraft in k-sets. The mathematical modeling used is called k-means, which is an efficient approximation when there are variations of certain characteristics among classes, but reducing a possible human subjectivity. Although the interest is to minimize possible subjectivities in aircraft classification, it is necessary to corroborate the mathematical analysis with the computational experience pratice. The method chosen to partition, k -means, is a process of easy development and computationally economic, therefore, feasible mainly because it does not require investment of very high costs [3]. This method is a generalization for the mean sample, so it is possible to analyze the asymptotic behavior pertinent to class characteristics and to determine some type of law based on the aircraft Big Data when analyzing the k of k-means for a large number.

In general, the method does not converge to a global optimal solution, although there may be special cases where this happens. However, as far as known, there is no feasible method by which a global solution is always obtained. Thus, it is guaranteed that at least one falibalistic result will be found, optimal local solution [3].

v.2, n.1, p.41-60, 2019.

METHODOLOGY

The mathematical modeling called k-means is developed as a possible tool for aircraft classification (taxonomy); then data acquisition, variables of the aircraft main characteristics, will be made to generate the database, therefore, the computational implementation will be done, for which the chosen programming language is Python. Thus, a validation analysis of the separated by the model groups can be done, which will make it possible to determine which direction the project should follow depending on the purpose of the aircraft.

DEVELOPMENT

Machine Learning

Types of Machine Learning

Machine learning can be done by feedbackings the input and output values, all of which is added to new data inclusion. In general, machine learning uses feasible mathematical models, that is, that they obtain solutions with a certain degree of accuracy, usually given by a Statistical Coefficient that varies between $[0, 1](\rightarrow 1)$. The greater the diversity of input values entered, the greater the gain of machine experience. Thus, the Correlation/Determination/Statistical Coefficient of the chosen adjustment function increases or decreases, this means that, a coefficient that has increased, has connections with major importance and, conversely, those that decreased the value of the coefficient are connections with minor importance. Thus, time is required for learning, depending on the type of problem to be solved. There are basically three types of machine learning: supervised, unsupervised and hybrid [4].

1. **Supervised:** the model receives an input data set and its corresponding output values, where adjustments are made in the mathematical-computational model until the difference between the output values generated by the model has a desirable value;
2. **Unsupervised:** the learning is constituted from input values in order to determine some data set characteristics, patterns in data, this, learning from self-adjusting model happens;
3. **Hybrid:** the supervised model is combined with the unsupervised, so that one layer can work with one type while the other works with the other type.

Dendrogramming

Dendrograms are pictographic representations of numerical elements, quantified by variables, belonging to a set or database (DB), which are visually contrasted as a function of the distances between the values of these same variables, dividing the set to which they belong in homogeneous clusters or subsets distinct from categories [5].

Dendrogramming as a computational technique takes place in the form of a phenotype matrix of dendrodistant elements $d_{i,j}$ which reduces the data set multidimensional space to a measure of distance between the elements. This measure, in the the Euclidean metric form, associates two elements $(e_i$ and $e_j)$ of this set, characterized by n variables, according to the next matrix object, Equation 1 [6]:

$$d_{i,j} = \sqrt{\sum_{v=1}^{n} (e_{i,v} - e_{j,v})^2} \tag{1}$$

Where:

$$\begin{cases} d_{i,j} = \text{Distance between } i \text{ and } j \\ v = \text{Label of variables } n \\ n = \text{Total number of elements} \\ e_{i,v} = \text{Variables } v \text{ of elements } i \\ e_{j,v} = \text{Variables } v \text{ of elements } j \end{cases}$$

The levels of (un)similarity or distance [5] between the sub-sets are observed through a dotted horizontal line or cut line made in the dendrogram: end of a line and beginning.

These abrupt change points identify the possible (un)groupings of the elements belonging to the set or database.

In the case of the dendrogram, Figure 1, it classifies six elements of the data set and indicates, at the level of similarity 0.50, two clusters: $\{1,2,3,4\}$ and $\{5,6\}$. If the dendrogram was cut by the dotted line down ($\rightarrow$ 1) or higher up ($\rightarrow$ 0), similarity levels would change, also changing the number of elements in clusters.

The dendrogramming method can be divided into hierarchical and not [7]. The first method is used when it is intended to identify the number of possible clusters while the non-hierarchical method requires the prior knowledge of the number of clusters.

As a multivariate statistical technique dendrogramming, has its epistemic origins in the biotaxonomy [8] which deals with the classification and ordering of biological organisms with similar vital characteristics [9].

From the point of view of modern phylogenetics, it is essential to dendrogram data related to DNA sequences as the basis for evolutionary studies and their physiological-aspectual changes in beings.

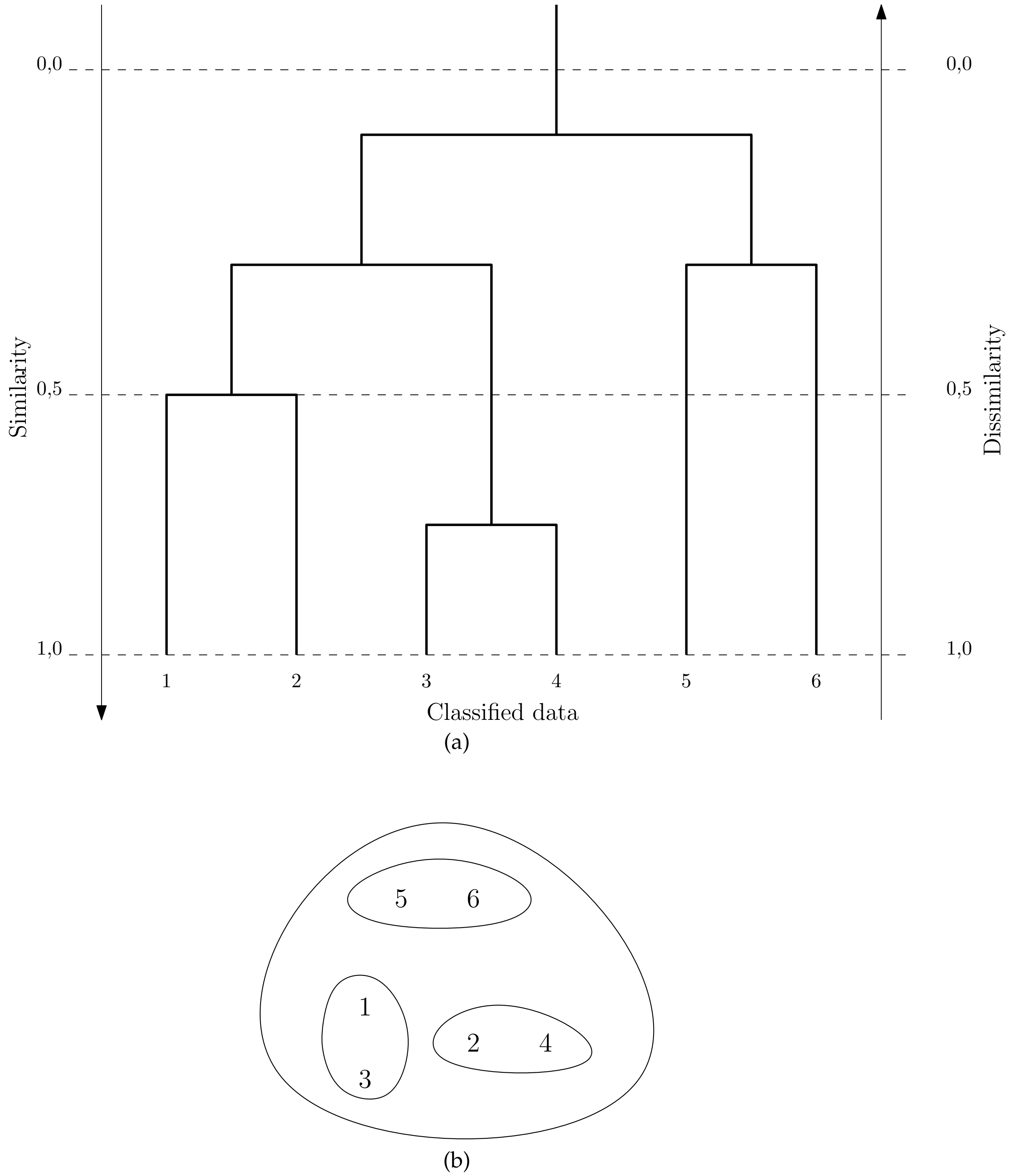

Figure 1. (a) Dendrogram: allegorical result of the Matrix Phenectic or Hierarchical Tree Dendrogram which indicate on the vertical axis levels of (un)similarity between the elements formed of sub-sets or clusters and in the horizontal axis the same elements are displayed.; (b) Grouping: result grouped according to the Venn Diagram.

v.2, n.1, p.41-60, 2019.

k-means

The *k*-means is an unsupervised clustering method that partitions data into *k*-groups where each group is nucleated by a medoid. Such partition, structured data mining from Big Data, implies a Voronoi Diagram, Figure 2.

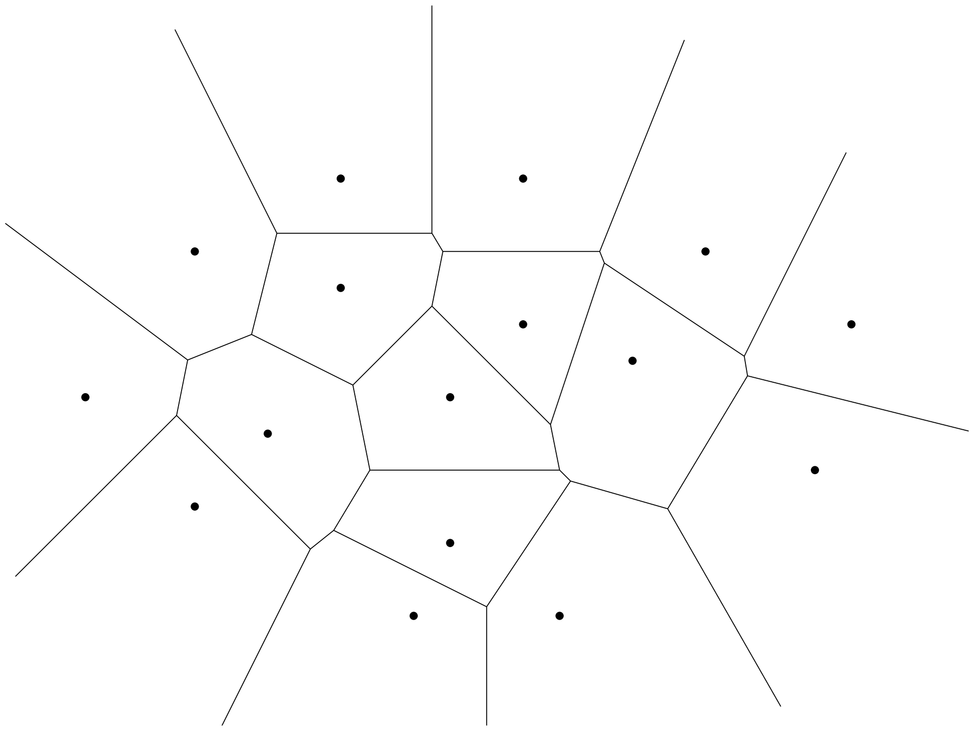

Figure 2. Example of a Voronoi Diagram.

Given an actual vector of a set of n observations $(x_1, x_2, x_3, \ldots, x_{n-2}, x_{n-1}, x_n)$, d-dimensional, the k-means method partitions the n observations into k-groups that minimize Equation 2.

$$WS = \sum_{k=1}^{K} \sum_{i=1}^{n} (e_{i,k} - \bar{m}_k)^2 \tag{2}$$

Where:

$$\begin{cases} WS = \text{Within Sum} \\ k = \text{Cluster } k \\ K = \text{Total number of clusters} \\ i = \text{Element } i \\ n = \text{Total number of elements} \\ \bar{m}_k = \text{Medoid of } k\text{-cluster} \end{cases}$$

Graphical Representation of the WS *Calculation*

For $K = 3$, three medoids ($\bar{m}_1, \bar{m}_2$ e $\bar{m}_3$) and nine elements ($n = 9$), where $9 = 3 + 4 + 2$, plots or elements of the respective elements contained in clusters $k = 1$, $k = 2$ and $k = 3$, results in k semiosis, Figure 3.

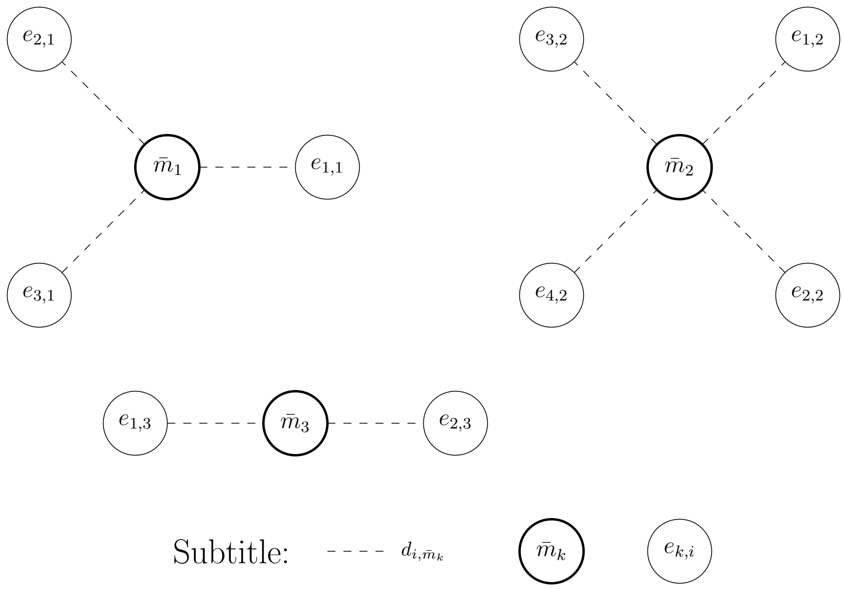

Figure 3. WS graphical representation.

Stopping Criteria

Iterative methods need some stopping criterion, otherwise they will run indefinitely ('infinitely'), not even obtaining one solution. When the criterion condition is satisfied, the best solution, until then is returned, even if it is not global optimal, in this way it is guaranteed that the method is applicable in feasible time instead of remaining infinitely in the search of an global optimal solution which, possibly, will not be found, also because, perhaps, there is no such convergence.

Therefore, one or more of the following items are adopted as a stop criterion:

1. The convergence of the mathematical method used. However, great care must be taken when choosing only this stopping criterion since it is difficult to guarantee that the method converges, that is, it is always recommended to use it with some other criterion;
2. A maximum number of iterations;
3. A maximum execution time;
4. A tolerance for the value of the objective function variation;
5. The number of times that the objective function value remains the same;
6. The number of times the value of the objective function variation remains the same;
7. The number of times that the value of the objective function variation remains the same in relation to an initial variation;

v.2, n.1, p.41-60, 2019.

8. Reach the number of combinatorial involved possibilities.

Elbow Data Calculation

The optimal number of k^*-clusters is indicated by the phenomenon called Elbow Data which is summarized in the analysis of the statistical value WS calculated at the end of the iterations.

The greater k-clusters number value, the smaller the WS value.

Results Associated with the Medoids As a Function of k

To better exemplify the Elbow Data phenomenon, one has the matrix 3 followed by tab. 1, from Figure 4 and 5, computational results for k ranging from 1 to 6, more details are contained in the description of Figure 4.

$$[BD]_{12\times 2} = \begin{bmatrix} 7 & 8 \\ 1 & 6 \\ 6 & 2 \\ 1 & 1 \\ 8 & 2 \\ 2 & 5 \\ 2 & 7 \\ 1 & 3 \\ 6 & 4 \\ 8 & 4 \\ 3 & 6 \\ 5 & 8 \end{bmatrix}_{12\times 2} \tag{3}$$

v.2, n.1, p.41-60, 2019.

Table 1. Results associated with medoids as a function of k [5].

k	(x, y)	(7, 8)	(1, 6)	(6, 2)	(1, 1)	(8, 2)	(2, 5)	(2, 7)	(1, 3)	(6, 4)	(8, 4)	(3, 6)	(5, 8)	Distance	WS
1	**(2, 5)**	5,83	1,41	5,00	4,12	6,70	0,00	2,00	2,23	4,12	6,08	1,41	4,24	15,41	25,87
2	**(7, 4)**	4,00	6,32	2,23	6,70	2,23	5,09	5,83	6,08	1,00	1,00	4,47	4,47	10,46	
1	**(1, 2)**	8,48	4,00	5,00	1,00	7,40	3,16	5,09	1,00	5,38	7,28	4,47	7,21	2,00	19,76
2	**(3, 6)**	4,47	2,00	5,00	5,38	6,40	1,41	1,41	3,60	3,60	5,38	0,00	2,82	12,11	
3	**(7, 3)**	5,00	6,70	1,41	6,32	1,41	5,38	6,40	6,00	1,41	1,41	5,00	5,38	5,65	
1	**(1, 2)**	8,48	4,00	5,00	1,00	7,00	3,16	5,09	1,00	5,38	7,28	4,47	7,21	2,00	13,64
2	**(2, 6)**	5,38	1,00	5,65	5,09	7,21	1,00	1,00	3,16	4,47	6,32	1,00	3,60	4,00	
3	**(6, 8)**	1,00	5,38	6,00	8,60	6,32	3,87	4,12	7,07	4,00	4,47	3,60	1,00	2,00	
4	**(7, 3)**	5,00	6,70	1,41	6,32	1,41	5,38	6,40	6,00	1,41	1,41	5,00	5,38	5,64	
1	**(1, 2)**	8,48	4,00	5,00	1,00	7,00	3,16	5,09	1,00	5,38	7,28	4,47	7,21	2,00	12,00
2	**(2, 6)**	5,38	1,00	5,65	5,09	7,21	1,00	1,00	3,16	4,47	6,32	1,00	3,60	4,00	
3	**(6, 8)**	1,00	5,38	6,00	8,60	6,32	3,87	4,12	7,07	4,00	4,47	3,60	1,00	2,00	
4	**(7, 2)**	6,00	7,21	1,00	6,08	1,00	5,83	7,07	6,08	2,23	2,23	5,65	6,32	2,00	
5	**(7, 4)**	4,00	6,32	2,23	6,70	2,23	5,19	5,83	6,08	1,00	1,00	4,47	4,47	2,00	
1	**(1, 2)**	8,48	4,00	5,00	1,00	7,00	3,16	5,09	1,00	5,38	7,28	4,47	7,21	2,00	10,00
2	**(2, 6)**	5,38	1,00	5,65	5,09	7,21	1,00	1,00	3,16	4,47	6,32	1,00	3,60	4,00	
3	**(5, 8)**	2,00	4,47	6,08	8,06	6,70	4,24	3,16	6,40	4,12	5,00	2,82	0,00	0,00	
4	**(6, 3)**	5,19	5,83	1,00	5,38	2,23	4,47	5,65	5,00	1,00	2,23	4,24	5,09	2,00	
5	**(7, 8)**	0,00	6,32	6,08	9,21	6,08	5,83	5,19	7,81	4,12	4,12	4,47	2,00	0,00	
6	**(8, 3)**	5,09	7,61	2,23	7,28	7,00	6,32	7,21	7,00	2,23	1,00	5,83	5,83	2,00	

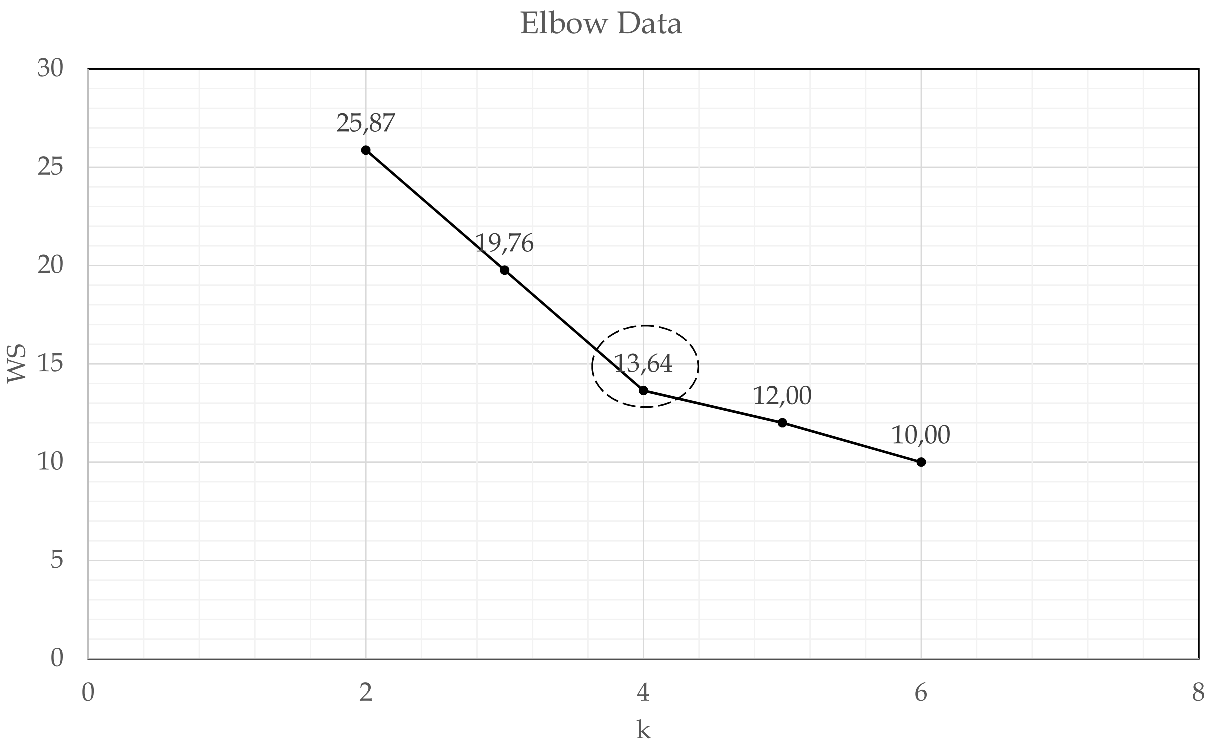

Figure 4. The suggested optimal number of clusters, $k^* = 4$, is indicated from the optimal Within Sum, $WS^* = 13.64$, circled value, hence from this point onwards, there are smaller and smaller variations for the WS value as k tends to an increasing number: the negative coefficient, tends to zero (horizontal), thus, WS^* is determined.

JOURNAL OF
PRODUCTION AND AUTOMATION

v.2, n.1, p.41-60, 2019.

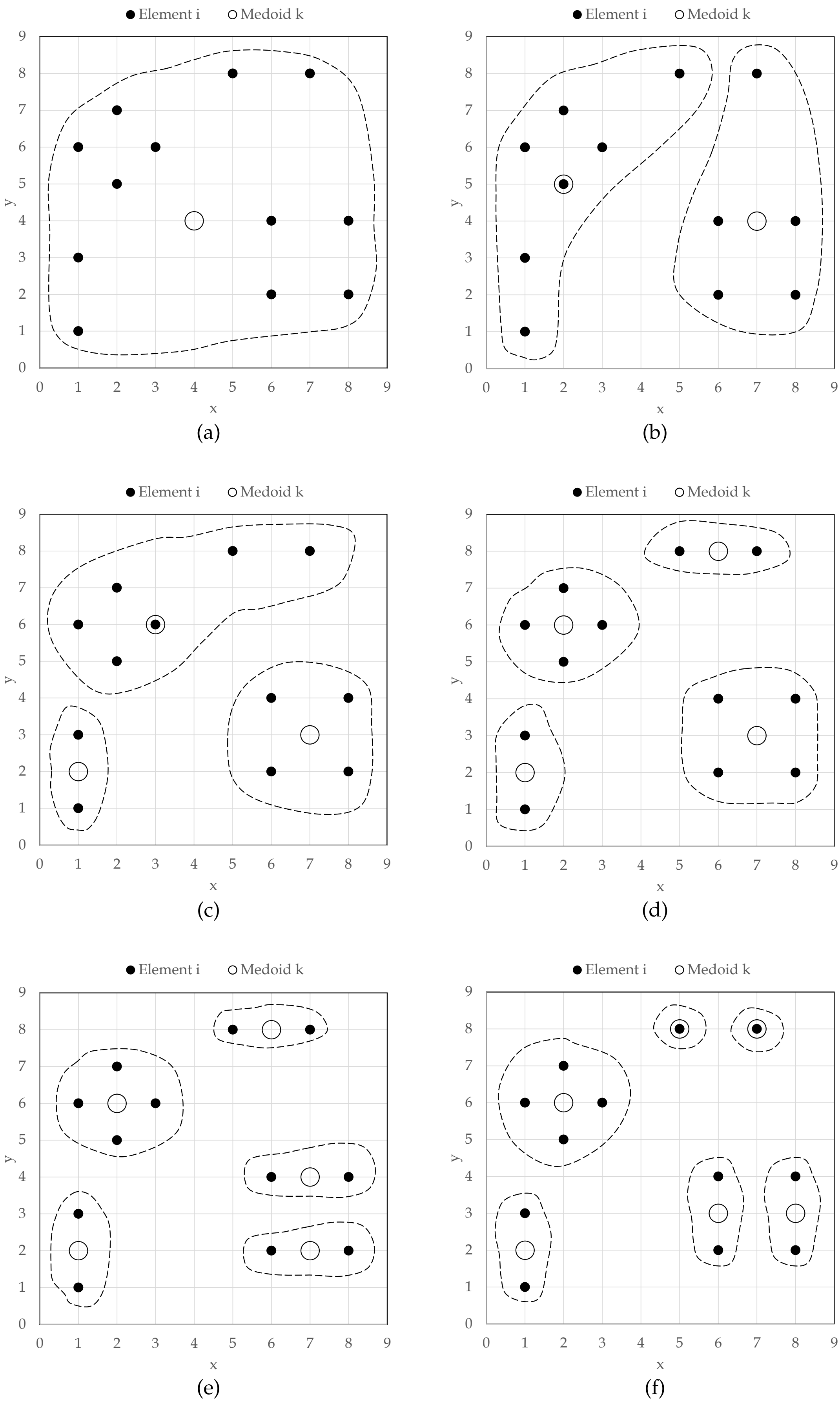

Figure 5. Partitions for: (a) $k = 1$; (b) $k = 2$; (c) $k = 3$; (d) $k = 4$; (e) $k = 5$; e, (f) $k = 6$.

Numerical Description

The Figure 6 shows computational implementation's flowchart.

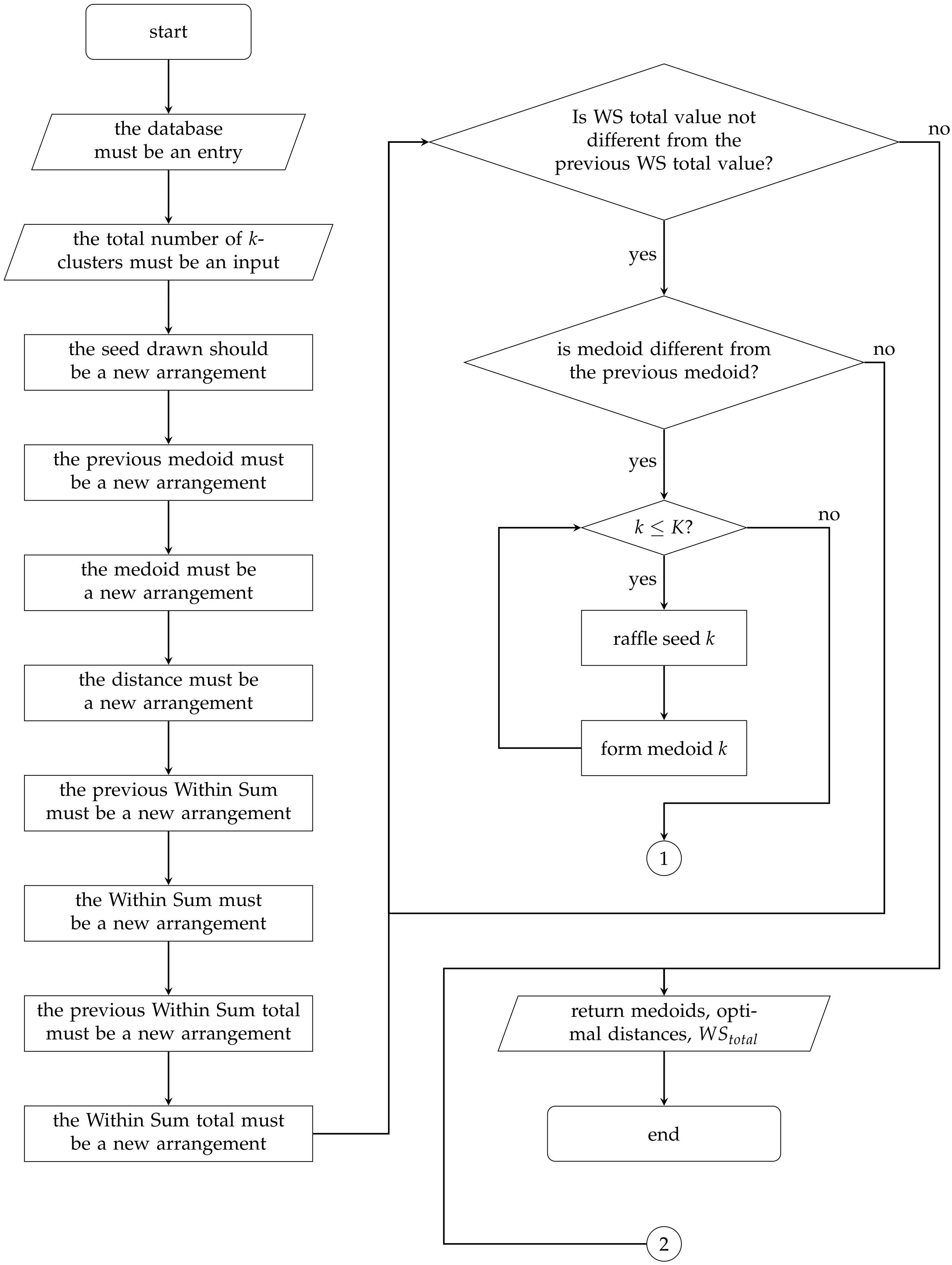

Figure 6. Flowchart of *k*-means

The Figure 7 shows shows the computational implementation flowchart's sequence.

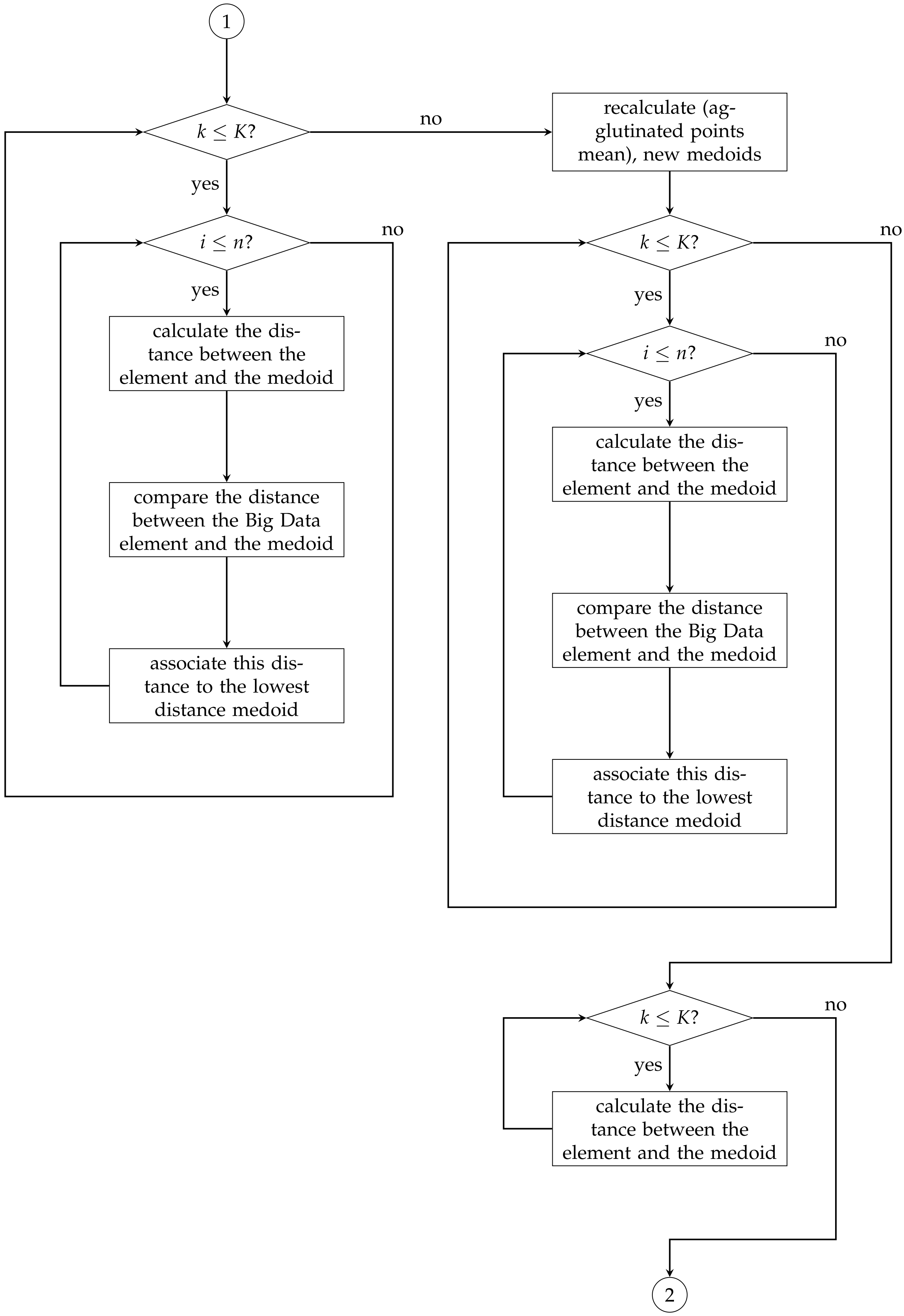

Figure 7. Flowchart of *k*-means (continued).

v.2, n.1, p.41-60, 2019.

Relevant Considerations to Development

Although Python has packages that, when invoked, have natively partitioning via k-means, the entire source code has been developed from the beginning. This is interesting because, in this way, both mathematics and computation involved are better known, and thus certain considerations and arbitrariness required by the problem can be made, which possibly are not or would not be possible from a "*black box*", since these are normally intended to suit all cases and therefore do not always have the parameters that the developer perceives to be necessary for a fine-tuning of the application that must be custom built. On the other hand, the programming and computing language is more and more known in a general way due to the knowledge that a computational, scientific application of this size demands. In addition, it would be quite tedious to analyze a third code, in case of some kind of problem, perhaps because "*black box*" was developed with a certain focus and/or for other reasons, regardless of the long time spent on the development. Obviously, fundamental resources are used from what is native to the language or package, such as generation of graphs and dendrograms, among other resources.

During the application development, some arbitrary others phenomenal details were defined and they that need to be highlighted separately:

1. Arbitrate: the adopted stop criterion is "Maximum number of iterations";
2. Phenomenon: to maintain the total number K of k-clusters, conditions are inserted so that the application does not draw a single seed more than once and, at the same time, it is guaranteed that a K number of medioids will be drawn;
3. Phenomenon: it does not make sense to generate results for a total number K equal to one, because it results in the database itself attracted by a single medoid, that is, there is no group partitioning;
4. Phenomenon: when the total number K of k-clusters tends to the total of elements n, the WS value tends to zero. However, it is not convenient to tend K to n, since this type of choice makes each of the elements themselves, in this case, the aircraft to be medoids. The problem is not in the fact that the elements are possible medoids, this may occur, but so that they will attract few or, in the most extreme case, only one element, itself. This results in inconsistent partitions, as it will result in the database itself where each element is a medoid and attracts itself. In addition to that, possibly causing the execution to enter in infinite loop. Arbitrated: a maximum K, less than or equal to $80\%n$ is considered;
5. The order of the database matters little;
6. The application generates file:
   a) Image of dendrogrammed graphs;
   b) Image of Elbow Data Chart.

FINDINGS AND DISCUSSION

As a result, it is expected the classification of aircraft in groups with certain characteristics or similarities, not only numerical but also ckassification via dendrograms. To do so the numerical values of the already developed aircraft numerical characteristics will be taken into account in order to reduce what can be understood as subjective. These results point to future project trends. The following are the considerations, discussions and results:

Considerations Pertaining to Executions

Beyond and in addition to the aforementioned considerations, to execute the application, the following considerations are used:

1. Arbitrated: the "maximum number of iterations" adopted is $1,000$ independent of the total value K of k-clusters;
2. Phenomenon: when the total number K of k-clusters tends to the total of elements n, the WS value tends to zero. However, it is not convenient to insert a K equals n, since this type of choice makes each of the elements themselves, in this case, the aircraft to be medoids. The problem is not in the fact that the elements are possible medoids, this may occur, but so that they will attract few or, in the most extreme case, only one element, itself. This results in inconsistent partitions, as it will result in the database itself where each element is a medoid and attracts itself. In addition to that, possibly causing the execution to enter infinite loop. Arbitrated: a maximum K, less than or equal to 80%n is considered;
3. It is not interesting to include the variable "Operation Entry Date" in the calculations;
4. The application generates file:
   a) Image of dendrogrammed graphs;
   b) Image of Elbow Data Chart.

Application Validation

To validate the application, the manually calculated databases are used, so the results for the comparison/validation were obtained. During the development of the application it the manual sorting of seeds was forced which assisted in debugging the application, since the manual calculation results were verified. The following are the computational outputs of the application, the intermediate values were omitted and the most relevant data is presented as optimal distance, WS optima, etc.:

Military Aircraft: Fighter

Variables

The variables used to describe a database of military aircraft are given by [10]:

1. Entry into service;
2. Engine type:
    a) reciprocating;
    b) turboprop;
    c) rocket;
    d) jet.
3. Engine power $[hp]$ or thrust $[N]$;
4. Wing span $[m]$;
5. Total length $[m]$;
6. Total height $[m]$;
7. Wing area (S) $[m^2]$;
8. Lower wing-sweepback angle $[degrees]$;
9. Higher wing-sweepback angle $[degrees]$;
10. Empty weigth $[kg]$;
11. Maximum takeoff weight (MTOW) $[kg]$;
12. Maximum level speed (VMO) $[\frac{km}{h}]$
13. Service ceiling $[m]$;
14. Range $[km]$;
15. Full armament payload coefficient $[kg]$;
16. Thrust-to-weight ratio $(\frac{T}{W})$;
17. Weight-to-area ratio $(\frac{W}{S})[\frac{kg}{m^2}]$.

Elbow Data Chart for $n = 40$ ***Military Aircraft and*** $K = 15$

Depending on the database used, you should run the application with a maximum K number higher or lower should be run. In this case, the application was executed up to a maximum value of $K = 15$. All this is done to try to find out the ideal number suggested by mathematics according to the Elbow Data Chart. The execution resulted in the Figure 8.

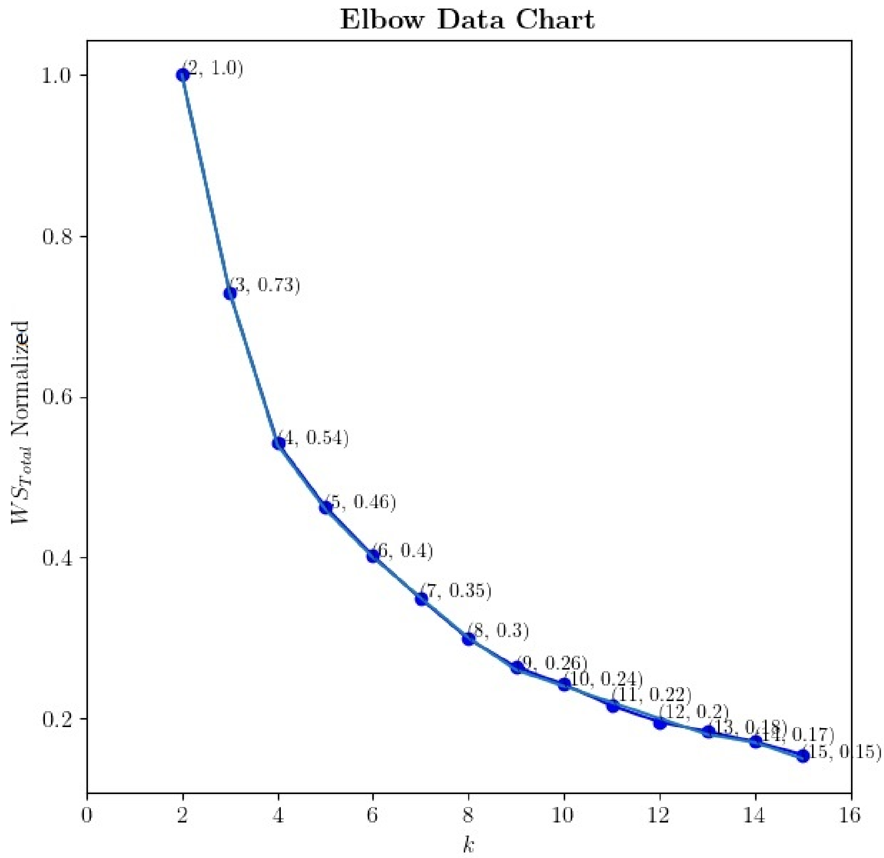

Figure 8. The suggested optimal number of clusters, $k^* = 10$, is indicated from the optimal Within Sum, $WS^* = 0.24$, normalized value, hence from this point on, decreasing variations to the WS value occur as k tends to an increasing number.

Thus, the next results are indicated for the K^* found from the Elbow Data Chart, Figure 8.

Discussions Concerning the Results of 40 ***Arbitrated Military Aircraft***

When analyzing the results for 40 military aircraft, Figure 8, arbitrated database, with some experience it is possible to realize that, generally, most of the aircraft contained in the database are projects considered good, great or even excellent. Here are some of the observations noted and classified in order of importance:

1. **Single aircraft:** The F/A-22 Raptor aircraft, the first fifth-generation fighter to come into operation, produced to dominate airspace until then, has a normalized optimum distance of 1.000 it, is necessarily unique when compared to the other 40, but still the MiG-31 M 'Foxhound' fighter is a very close project,

numerically, and therefore the best fighter to go into combat with the F/A-22 Raptor. This type of information is extremely important in a confrontational scenario, not only for these two fighters, but also for the other possible comparasions. To succeed in a confrontation, it is imminent to know which fighter should take off to combat the enemy aircraft, which is why radars and satellites identify, through reconnaissance of images, in addition to other resources, the aircraft at all times. It is also known that it is not appropriate to combat a very small aircraft with a larger one, due to, among other things, the operational cost and different speeds of stall;

2. **Comparison with national aircraft:** Embraer EMB-314 aircraft (Super Tucano) was intentionally inserted into the database to further test the mathematical-computational application and extract some information or knowledge. Note that this one remains alone in its "group", in this case, 1, because it is a fighter that, in fact, as expected, does not fit the other fighters because of the type of mission for which it was designed. This is yet another confirmation that everything that has been done is correct.

Dendrogram For $n = 40$ *Arbitrated Military Aircraft*

The dendrogram for the previous results, that is, for $k^* = 10$ groups is represented in Figure 9.

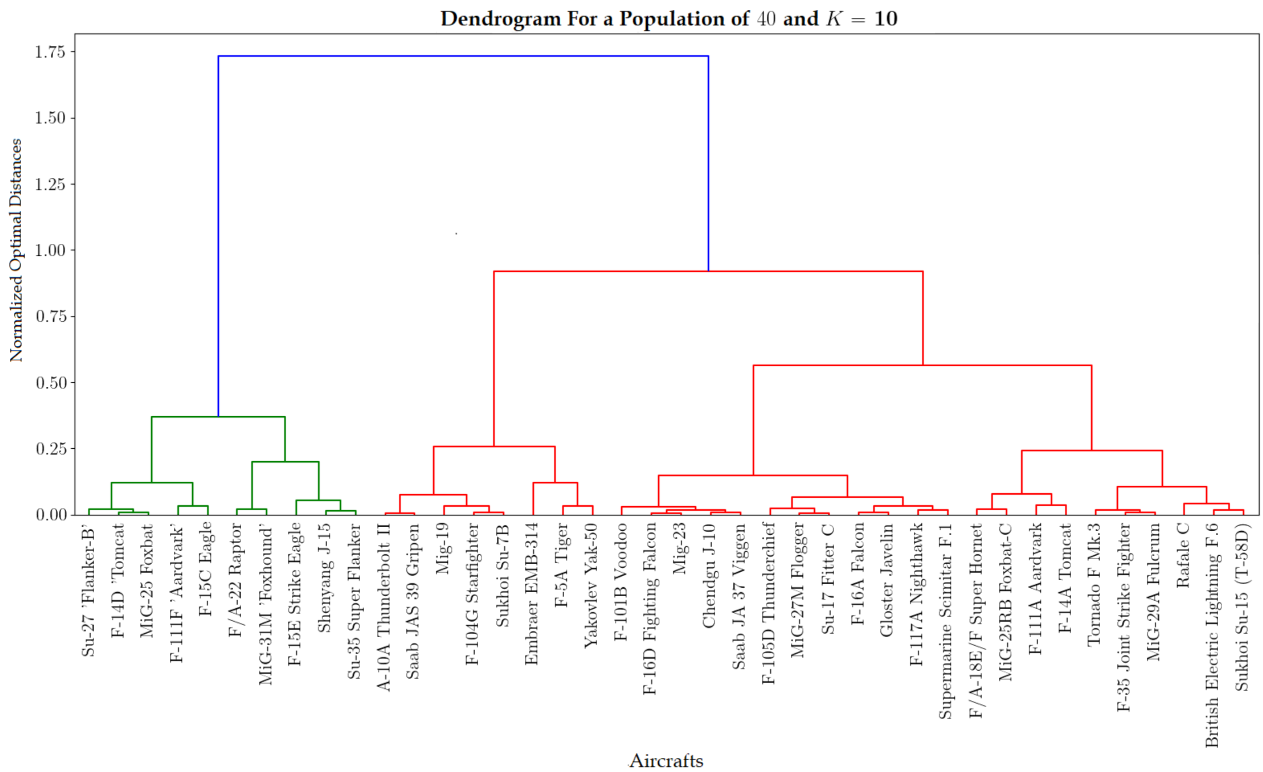

Figure 9. Dedrogram computational results for 40 arbitrated aircraft and $K = 10$.

v.2, n.1, p.41-60, 2019.

CONCLUSIONS

Developing the study and the activities, it was possible to observe that, the project went in the right direction, based on historical solutions, the chosen mathematics is solution to the problem of partitioning the aircraft. The presented method application was generically developed which serves as a mathematical-computational tool to classify other databases.

One of the major problems of the aerospace industry, in particular the construction of military aircraft, is the expense in theses vehicles pre-design. It is common to invest huge expenses and a great deal of the design engineers time to discover that the focused project is simply unsuitable for a mission or the market. In some cases, all of this is discovered after wasting all the money and time, which can lead to the company's bankruptcy. Therefore, the *k*-means, Big Data aerotaxonometric technique allows, as confirmed in this work, a better initial orientation for a possible successful project, in terms of its beginning, middle and end; adjusted according to the prevailing preferences of the aeronautical market. Another benefit is the evolution of the current engineering towards of the future pointed out by Big Data that, besides prescribing working time minimization it, also points to a reduction in scale of the companies economic-financial effort, in a general way.

The *k*-means deployed in dendrogramming allows through its dendros, the allocation of a future project, by similarities of data to be associated pictographically with all the other projects, already realized, with success or failure from the commercial point of view, the functional or both. This aerotaxonomic classification enables the aerospace engineer, by hand and instantaneously, to have a well-founded direction for his future aircraft whose possible modification in this or that design characteristic, allocates to this or that dendro, in that one or in a previously designed set of aircraft. Thus, during the development of the project, it can be verified at any stage, if it is within a pre-stipulated classification and, if it holds to an unwanted, unwanted, it is possible to make the necessary technical modifications so that the project does not take another direction and its budget, most of times in *deficit*, is not consumed. Another point that all this mathematics highlights, are possible similarities between aircraft never compared before due to the large number of projects and variables, among other details. Not only that, but also brings the possibility of designing a more cost-effective product, therefore becoming more competitive.

It has also been found that the *k*-means numerically functions as a kind of planetary system where the medoids are the stars and the elements are planets that are gravitated by an attraction force produced by these bodies of greater mass, the medoids. Consequently, the model, by varying the medoids position (stars) at each iteration, converges to the optimal point (of minimal action) and, from there, the position of the medoid does not change, since the system balance point was found, all without the need to execute "endless" iterations due to the quick method convergence. Therefore, the mathematical-computational iterative adopted method eliminates the natural misunderstandings of previous iterations, once the convergence adopted criterion and, above all, the developed mathematics, causes the equilibrium point to be reached. That is, an element mistakenly allocated in a medoid that is not its ideal, will be relocated with each step towards its ideal medoid, until stabilizing in this one.

v.2, n.1, p.41-60, 2019.

Concerning commercial domain, and especially military, this mathematics, current and based on large amounts of data (Big Data), is fundamentally instrumental in the analysis, comparison and development of projects. It is no wonder that the companies in this branch of engineering are seeking more and more professionals with this type of knowledge, who know not only specific solutions, but also make it possible to see the market in a broad, above all, circumstantial way.

Analyzing the results we come to certain conclusions that only a very experienced person would know. However, the *k*-means brings the possibility that a person who is not an expert in this type of work can extract knowledge through the information (database), but also for those who are experted to perceive certain nuances that only a deeper analysis is able to highlight, far beyond what has already been done. It is worth remembering that the discussions carried out in this work, are not limited here, there is more to be discussed about all of this, thanks to the new mathematical methods that have been emerging in the last decades which bring innumerable possibilities. Another detail is that the already known information can be confirmed as it is quoted and compared throughout the text, it is worth remembering the case of F/A-22 Raptor previously discussed.

Mathematics is adequate and has shown, despite being falibalist, with a high degree of accuracy, an excellent method for partitioning databases; in fact, the obtained results in addition to other possible analyzes that can be made serve as the basis for new projects and not only that, but the whole study can be used in other areas of knowledge or also several analyzes within the same work can be made, in this case, as if it were a kind of decision tree. When executing the partitioning of the database, a group from the dendros is chosen, in which a new classification is made, which will result in another dendrogram, and thus, one can extract much deeper and specific results regarding the aircraft can be extracted. Likewise, when choosing a certain group as the basis for a new project, this same analysis may suggest the separation of aircraft subsystems (aerodynamic surfaces, propulsion systems, control, avionics, communication and structural components, among others) chosen, then the new project subsystems, with the best value from that initially adopted aircraft group, can be chosen.

The developed work it possible to follow market trends, as it is based, as seen, on database of what has already been produced, or rather, it is necessary to do the analysis with some frequency to follow such modulations and, this will in fact lead to what the market demands in financial terms and at the same time with regard to its mission. This allows the company to be constantly renewed in terms of market trends and requirements, a fundamental requirement for companies to remain in business competition as well as for a nation to maintain its air sovereignty.

Finally, it was possible to perceive in the *k*-means, not only these, but also other extremely comprehensive and eminently functional solutions for engineering, but with a different style of mathematics that contrasts with the classical ways. Grounded in the aircraft DNA (set of variables), by distancing itself more and more from subjectivity, mathematics traced the likelihoods between them, that is, the non-absolute, however true similarity. This method, interchangeable, is versatile, liquid enough to keep up with societal perspectives, and therefore carries within it a real probability of having the future engineering mathematics in hand, in the case, the present.

v.2, n.1, p.41-60, 2019.

REFERENCES

[1] Linnaeus, C. v. Systema naturae per regna tria naturae. secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis, 1758.

[2] Bauman, Z.. Liquid Modernity. Polity Press, 2000. ISBN 0745624103.

[3] MacQueen, J. Methods for classification and analysis of multivariate observations. volume 28: (1967), páginas 250–274. ISSN 0883-7252. doi:10.1002/jae.1272.

[4] Russell, S. e Norvig, P. Artificial Intelligence: A Modern Approach. Prentice Hall, 2 edição, 2002. ISBN 978-0137903955.

[5] Ricieri, A. P. Curso de big data de lousa. Prandiano - Museu da Matemática, São Paulo, 2018.

[6] Berk, K. N., Khattree, R. e Naik, D. Applied Multivariate Statistics with SAS Software, volume 54, 2000. doi:10.2307/2686040.

[7] Anderberg, M. R. Cluster analysis for applications, 1973.

[8] Manktelow, M. History of taxonomy. Uppsala University, 2010.

[9] An update of the angiosperm phylogeny group classification for the orders and families of the flowering plants. Botanical Journal of The Linnean Society, volume 161: (2009), página 105. ISSN 0024-4074. doi:10.1111/j.1095-8339.2009.00996.x.

[10] Silva, B. M., Fregnani, J. A. T. G. e Magalhães, P. E. C. S. Frontiers in aerospace science: conceptual design of green transport airplanes. Bentham Book, 2018.

CONTACT INFORMATION

Eden Denis Ferreira da Silva Lopes Santos (corresponding author)
edendenis@gmail.com

Carlos Alberto Rocha Pimentel
carlos.pimentel@ufabc.edu.br

Aguinaldo Prandini Ricieri
contato@prandiano.com.br