# APPLICATION OF ITEM RESPONSE THEORY IN ASSESSING THE RELEVANCE OF ELEMENTS IN CLUSTERS

# APLICAÇÃO DA TEORIA DE RESPOSTA AO ITEM NA AVALIAÇÃO DA RELEVÂNCIA DE ELEMENTOS EM GRUPOS

# APLICACIÓN DE LA TEORÍA DE LA RESPUESTA AL ÍTEM EN LA EVALUACIÓN DE LA RELEVANCIA DE LOS ELEMENTOS DE LAS AGRUPACIONES

**Paulo Fernando Leite-Filho**[1]

**ABSTRACT:** The objective of this study is to apply Item Response Theory (IRT) to determine the difficulties in assigning elements to their respective groups of 38 clustering algorithms in 11 datasets. Using the cognitive scale based on the Difficulty b-Parameter of the IRT Logistic Parameter Models (LP), results were obtained indicating significant changes in the relevance of elements in balanced and unbalanced groups, where the level of accuracy in assigning elements in the difficult range was predominant for the clustering algorithms. The use of the psychometric scale during grouping assignments in adverse situations increases the reliability of decision support systems, with potential use by professionals in critical areas.

**Keywords:** Machine Learning. Cluster. Item Response Theory.

**RESUMO:** O objetivo deste estudo é aplicar a Teoria de Resposta ao Item (TRI) para determinar as dificuldades na atribuição de elementos a seus respectivos grupos de 38 algoritmos de agrupamento em 11 conjuntos de dados. Usando a escala cognitiva baseada no parâmetro b de dificuldade dos modelos de parâmetros logísticos (LP) da TRI, foram obtidos resultados que indicam mudanças significativas na relevância dos elementos em grupos equilibrados e desequilibrados, onde o nível de precisão na atribuição de elementos na faixa difícil foi predominante para os algoritmos de agrupamento. O uso da escala psicométrica durante as atribuições de agrupamento em situações adversas aumenta a confiabilidade dos sistemas de apoio à decisão, com potencial de uso por profissionais de áreas críticas.

**Palavras-chave:** Aprendizado de máquina. Cluster. Teoria de resposta ao item.

**RESUMEN:** El objetivo de este estudio es aplicar la Teoría de Respuesta al Ítem (TRI) para determinar las dificultades en la asignación de elementos a sus respectivos grupos de 38 algoritmos de agrupación en 11 conjuntos de datos. Utilizando la escala cognitiva basada en el Parámetro b de Dificultad de los Modelos de Parámetros Logísticos (LP) de la TRI, se obtuvieron resultados que indican cambios significativos en la relevancia de los elementos en grupos equilibrados y desequilibrados, donde el nivel de precisión en la asignación de elementos en el rango de dificultad fue predominante para los algoritmos de agrupamiento. El uso de la escala psicométrica durante la asignación de agrupaciones en situaciones adversas aumenta la fiabilidad de los sistemas de apoyo a la toma de decisiones, con un uso potencial por parte de los profesionales en áreas críticas.

**Palabras clave:** Aprendizaje automático. Cluster. Teoría de la respuesta al ítem.

[1] Doutorado
Universidade Federal de Pernambuco - UFPE

## INTRODUTION

Algorithms learn in different ways to solve problems with specific objectives, enabling greater algorithm efficiency in different types of databases [1]. Clustering algorithms use the unsupervised learning methodology and aim to group elements into clusters based on their similarity, aiming to assign a cluster label to each sample [2]. Clustering algorithms can be penalized in the formation of clusters by factors such as prior determination of the number of classes in clusters, data sampling problems, dimensionality variation, and class imbalance. There are related factors such as sample characteristics, providing limitations to the algorithms, compromising the formation of clusters, and causing indeterminations in the belonging of the elements in the respective groups [3][4]. Assessing the difficulty in representing the importance of the elements contained in the clusters in relation to their respective pertinence in the groups generated by clustering is the objective of this work.

## CLUSTERING ALGORITHMS AND THE MEMBERSHIP OF ELEMENTS IN GROUPS

The following challenges are attributed to clustering by algorithms: i) representation of the importance of the elements contained in the clusters in relation to their respective membership in the groups; ii) interpretation of the quality of the cluster, based on the use of the distances of the cluster elements, which is difficult to interpret due to the generalization of the grouping formed in relation to the diagnostic label assigned by the experts; iii) the loss of information due to the spatial geolocation of the element in the cluster; and iv) inadequacies due to the loss of information in the multiple correspondence analysis [5]. Clustering algorithms use data to construct data groups to solve specific problems in several complex decision support systems, commonly applied in diagnostics of health datasets from emergency rooms [5]; analysis of gene expression data to identify expression patterns related to genetics [6]; for diagnosis of hereditary defects, segmentation of medical images, detection of social communities, anomaly detection, class extraction, analysis of consumer behavior [2] and other applications.

## CLUSTERING ALGORITHMS AND IRT

Item Response Theory (IRT) is a psychometric method in Psychology that allows for reliability assessments of measurement instruments, such as academic assessments, questionnaires, and similar instruments [7]. IRT can also be seen as a statistical method used in assessments and questionnaires to identify respondents' skills, and which also makes it possible to individually classify each item into different levels of difficulty in the responses obtained [8]. IRT has been considered a solution for selecting records considered difficult or easy. Identifying the level of significance of each item in relation to the set allows for the most optimized decision-making for each application in their respective areas. When associating IRT with grouping and classifications, it is understood that the instances correspond to the items and the clustering algorithms correspond to the respondents. This type of association has been referenced in the literature in experiments with supervised Machine Learning, as in the work of [9], in this way, IRT can be applied in the evaluation of the relevance of elements in groups and how clustering algorithms create the groups.

## ITEM RESPONSE THEORY

Item Response Theory is a statistical method based on probabilistic models to determine the relationship between test items and individuals' ability, which defines the probability of an examinee's response to a test item as a function of latent ability of the examinee and the characteristics of the item. For this, a method was developed that calculates the skills of a student, and presents a parameter. For the difficulty of the test items, calculated by Equation 1 [10].

$$P(U_i = 1|\theta_j) = c_{i+} \frac{1 - c_i}{1 + \exp(-a_i\,(\theta_j - b_j))} \quad (1)$$

$$c_{i+} \frac{1 - c_i}{1 + \exp(-a_i\,(\theta_j - b_j))} \quad (1)$$

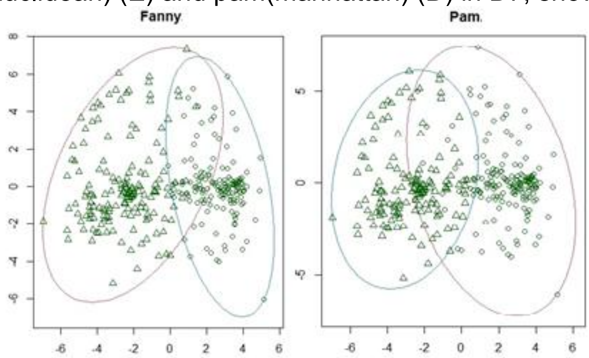
Where: P(U_i=1|□_j): The function of the parameters for the effects of the classifier skill and the properties of the item; i: The index of the sorted item; j: The index of the classifier and θ: Classifier skill. The following variables correspond to the IRT Logistic Parameters: ai: Discriminating item; bi: Item Difficulty and cj: Probability of the classifier getting it right at random (Guessing). These IRT parameters are present in the three Logistic Models – LM, these models combine the parameters of Discrimination (a), Difficulty (b) and Probability of hitting at random (c). The Logistic Parameters (LP) are represented in Equation 1 for each of the LM. Each model has an important role in the evaluation of the item in relation to the set to be classified by the algorithms [5][12]. For LP Models, the values of each IRT parameter can vary in 1LP (a = 1, b = [-4,+4], c = 0), 2LP (a = [-4,+4], b = [-4,+4], c = 0) and 3LP (a = [-4,+4], b = [-4,+4], c = [-4,+4]). However, in a practical application, the algorithms are only evaluated in the number of hits, unlike an application in a critical system, where issues of complexity are evaluated when making decisions at critical moments, detecting anomalies [9] or analyzing financial [6].

PROBLEMS IN FORMING CLUSTERS

The factors that influence the performance of clustering algorithms are the objectives of the algorithms, sample size problems, class balance and boundary problems. These factors can be assessed in their formation of groups of distinct elements by two types of cluster validation measures, which are based on criteria for the formation of elements external and internal to the clusters [13][14].

Figure 1 shows two graphs corresponding to different clustering algorithms in the same data set, where it is possible to observe that each algorithm created its respective partitions using different techniques used in each clustering algorithm, as well as by the use of different dissimilarities, thus it is possible to observe that the elements can be assigned to different groups, depending on the clustering algorithm and the group formation method applied by it. Where it is possible to observe the conflict of correctly identifying the elements in the respective clusters, considering the boundary area of the elements correctly identified in the respective groups and those identified incorrectly (i.e. the noise).

Figure 1 – Algorithms fanny(SqEuclidean) (E) and pam(manhattan) (D) in D7, showing the different groups formed.




Source: Author

For a more objective assessment, references to hits and misses in the assignments of elements in the respective groups are used. Since clustering algorithms do not use previously assigned labels (i.e. unsupervised), it is necessary to determine a comparative method between the assignments of elements in the groups and the significance of these assignments.

**METHOD**

The algorithms implemented in the experiments performed in this work were coded in R language version 4.2.2. The continuous data of the datasets were transformed into dichotomous data using the Simulate Response Patterns technique of the R “mirt” package for IRT applications [15]. For the experiments, restrictions were considered in the experiments to evaluate the difficulties calculated by the Difficulty Parameters b of the three IRT LP Models, in the assignments of the elements in the respective groups performed by the clustering algorithms. The 34 algorithms used in the experiments of this work are described below. The algorithms: Fuzzy Analysis Clustering (fanny); Partitioning Around Medoids (pam); Clustering Large Application (clara); Self-Organization Tree Algorithm (sota); K-means; Divisive Analysis (Diana); KMeans_arma and KMeans_rcpp; Mini Batch Kmeans; Fuzzy C-means; Cluster Medoids; Gustafson-Kessel; Multi-objective Optimization for Collecting Cluster Alternatives (MOCCA).

Dissimilarities were the characteristics used in each type of clustering algorithm to differentiate the methods for determining the distances of elements in clusters, providing a broader view of the formation of clusters formed by each specific algorithm. For the fanny, pam and clara, and diana algorithms, the Euclidean, Manhattan, Square Euclidean and Jaccard dissimilarity measures were applied. For the K-means algorithm, centroid positioning assignments were used, using the Hartigan-Wong, Lloyd, Forgy and MacQueen methods, tested in the work of [16]. Centroid positioning assignments were also used for K-means. For the Mini Batch Kmeans algorithms, the initialized positioning of the centroid was used as a characteristic of these algorithms as an objective, using centroid initialization techniques represented by Kmeans++, Random, Optimized Initialization and Qualitative Initialization [17]. For the Fuzzy C-means algorithms, elements contained in multiple clusters were evaluated, as in the work of [18]. For the cluster medoids algorithms, the objective was to evaluate the elements defined as centroids or medoids in the clusters, providing a reference in the distances calculated by the dissimilarities of the elements of the groups obtained, a method used in the work of [19] to evaluate the formation of clusters of data obtained from the internet. The Cluster Medoids and Gustafson-Kessel algorithms are more current and aimed at a comparison in the evolution with the algorithms of the other groups. These algorithms were applied in 11 datasets.

The datasets were normalized with mean equal to zero and variance equal to 1, these datasets were organized from D1 to D11, and represented in unbalanced and balanced groups using the SMOTE method, where: D1 (Echocardiogram: Unbalanced (Class 0 = 69 (67.20%), Class 1 = 42 (32.80%)), Balanced (Class 0 = 172 (49.57%), Class 1 = 175 (50.43%))); D2 (Heart-stat log; Unbalanced (Class 0 = 102 (44.40%), Class 1 = 128 (55.60%)), Balanced Class 0 = 260 (56.52%), Class 1 = 200 (43.48%))); D3 (Ionosphere; Unbalanced (Class 0 = 206 (64.10%), Class 1 = 115 (35.90%)), Balanced Class 0 = 444 (51.51%), Class 1 = 418 (48.49%))); D4 (Parkison; Unbalanced (Class 0 = 140 (75.40%), Class 1 = 45 (24.60%)), Balanced (Class 0 = 273 (48.49%), Class 1 = 290 (51.51%))); D5 (Blood Transfusion; Unbalanced (Class 0 = 570 (76.20%), Class 1 = 178 (23.80%)), Balanced (Class 0 = 1140 (51.63%), Class 1 = 1068 (48.37%))); D6 (Diabetes; Unbalanced (Class 0 = 268 (34.90%), Class 1 = 500 (65.10%)), Balanced (Class 0 = 1072

(51.74%), Class 1 = 1000 (48.26%))); D7 (Madelon; Unbalanced (Class 0 = 1300 (50.00%), Class 1 = 1300 (50.00%)), Balanced (Class 0 = 1300 (50.00%), Class 1 = 1300 (50.00%))); D8 (Vinnie; Unbalanced (Class 0 = 185 (48.68%), Class 1 = 195 (51.32%)), Balanced (Class 0 = 390 (51.32%), Class 1 = 370 (48.68%))); D9 (Glass; Unbalanced (Class 0 = 138 (64.49%), Class 1 = 76 (35.51%)), Balanced (Class 0 = 276 (47.59%), Class 1 = 304 (52.41%))); D10 (Banana (18.87% of original); Unbalanced (Class 0 = 1000 (50.00%), Class 1 = 1000 (50.00%)), Balanced (Class 0 = 1000 (50.00%), Class 1 = 1000 (50.00%))) and D11 (Breast cancer; Unbalanced (Class 0 = 317 (100.00%), Class 1 = 366 (100.00%)), Balanced (Class 0 = 317 (34.99%), Class 1 = 366 (65.01%))), Retrieved from (https://github.com/nandomp/IRT4ML) and (https://archive.ics.uci.edu/).

All datasets are binary and with the elements pre-defined in their respective classes. In a semi-supervised manner, the respective groups will be considered by the classes of the datasets informed, thus providing validation for the assignments of the elements in the respective groups by the clustering algorithms. In the literature, it is possible to assess the formation of groups using cluster validation indices [20], which can be internal and external. Internal indices such as Silhouette [21], Dunns [22], Davies-Bouldin [23], and PBM [24]; and the external indices Foulkes-Mallows [22], Rand [25], and Jaccard [25]. In this work, a psychometric scale based on IRT will be used to assess the levels of difficulty.

## PSYCHOMETRIC SCALE BASED ON IRT

To calculate the different levels of difficulty by IRT, six ranges were categorized for the reference of the Difficulty Levels (DL) of the values of the Difficulty Parameter (b) in levels of complexity based on IRT, these levels are: from [- infinity to -4] is extremely easy (XE); from [-4,-2] is very easy (VE); from [-2,0] is easy (E); from [0,+2] is difficult (H); from [+2,+4] is very difficult (VH); and from [+4,+ infinity] is extremely difficult (XH).

These ranges were adapted from the work of [26], where the levels of the Difficulty b-parameter were divided as follows: MFC (Very easy and correctly classified) [-4,-2], FC (Easy and correctly classified) [-2.0], MDC (Very difficult and correctly classified) [0,+2], DC (Difficult and correctly classified) [+2,+4], MFE (Very easy and incorrectly classified) [-4,-2], FE (Easy and incorrectly classified) [-2.0], MDE (Very Difficult and incorrectly classified) [0,+2], DE (Difficult and incorrectly classified) [+2,+4]. The ranges smaller than -4 and greater than +4 were adjusted to adapt values that extrapolate the possible results recorded.

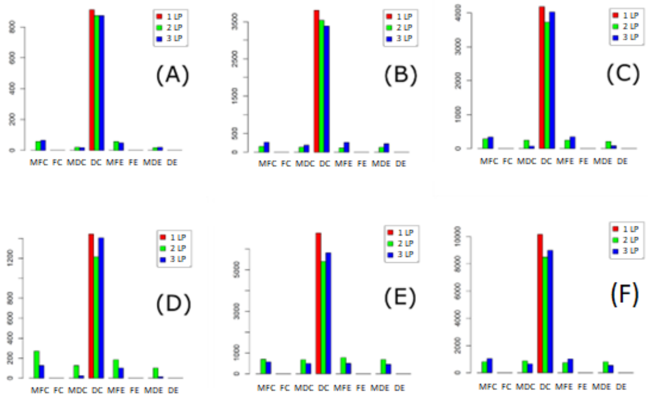
## RESULTS AND DISCUSSION

The use of the IRT Difficulty b-Parameter determination method makes it possible to identify which items can be identified with assignments to the groups in a complex way. The Difficulty b-Parameter variable is present in the three IRT MLs [27]. The X axis shows the correct categories (MFC, FC, MDC, and DC) or error (MFE, FE, MDE, and DE), and the Y axis of the graphs shows the sum of classified items from all datasets.

It is possible to observe that when performing oversampling using SMOTE, the number of errors is reduced and the level of difficulty is also reduced in relation to the three LP Models.

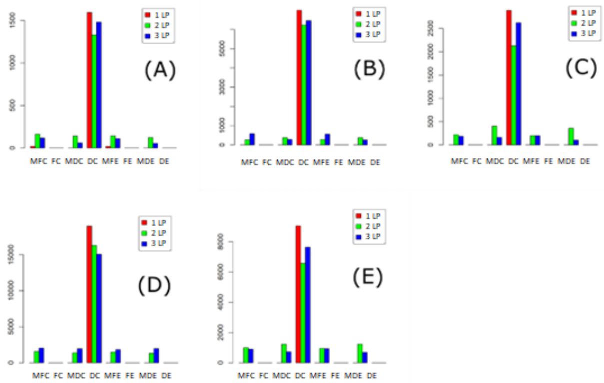
Figures (2-3) shows the results of group 1 with 50% undersampling of the minority class in each dataset, considering the intervals mentioned in the psychometric scale defined in the previous paragraph.

Figure 2 – Graphs of the values obtained by the psychometric scale from datasets 1(A) to dataset 6(F) of group 1.




Source: Author.

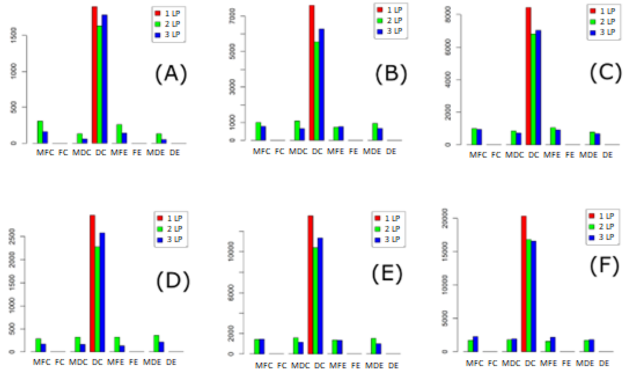
Figure 3 – Graphs of the values obtained by the psychometric scale from datasets 7 (A) to dataset 11 (E) of group 1.




Source: Author.

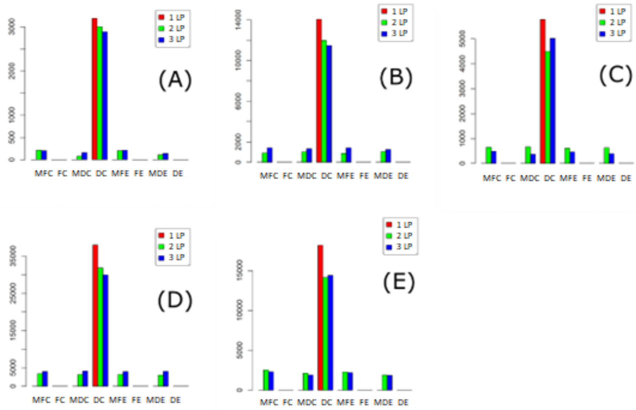
Figures (4-5) shows the results of group 2 with 100% subsampling of the minority class in each dataset, considering the intervals mentioned in the psychometric scale defined in the previous paragraph.

Figure 4 – Graphs of the values obtained by the psychometric scale from datasets 1 (A) to dataset 6 (F) of group 2.




Source: Author.

Figure 5 – Graphs of the values obtained by the psychometric scale from datasets 7 (A) to dataset 11 (F) of group 2.




Source: Author.

In the work of [8] and [28], the complexity scales are related to the IRT Difficulty b-Parameter and it is demonstrated that errors in classifications can occur due to data complexity. In the experiments of this work, the unsupervised algorithms were evaluated with imbalanced datasets and balanced with subsampling.

In the work of [8], mostly supervised algorithms and balanced datasets were evaluated, and part of the datasets tested in the experiments were used in the tests of this section, where the algorithms based on k-means again obtained good results.

In Figures (2-5), it is possible to observe that there is a concentration of items in grouping assignments in relation to the previously known classifications in the Difficult and Correct (DC) category in relation to the other categories.

However, by adding more elements from group 1 to group 2, it is possible to observe a distribution among the other categories, as can be seen in Figure 2(A) in relation to Figure 4(A). This behavior can also be observed in the other datasets of the experiments carried out in this work.

Evaluating the behavior of algorithms in relation to the variation of element concentrations in relation to the test groups is a practice known as minority oversampling (ROS) and random majority undersampling (RUS), as well as artificial data generation techniques such as SMOTE.

Group imbalance is a factor that penalizes the performance of clustering algorithms, and as reported in the previous paragraph, balancing by subsampling the majority group is a promising method and has been tested in the experiments in this section.

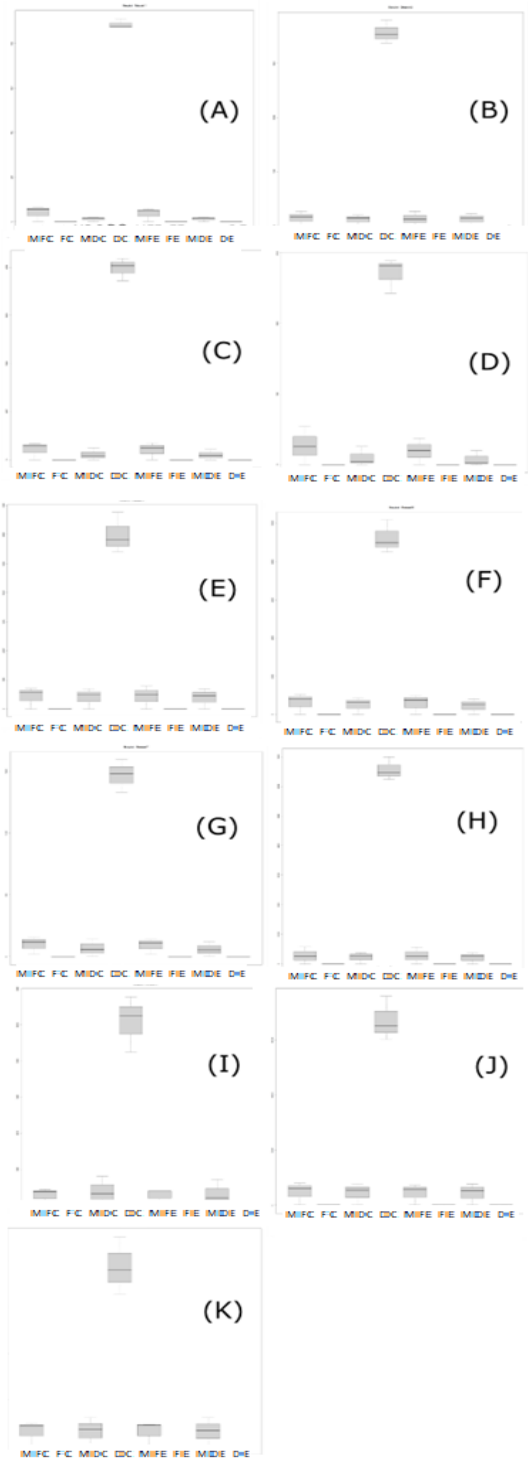
Applying subsampling to solutions involving clustering algorithms presents limitations in important information with significant amounts lost during this process. Cluster-based subsampling techniques solve problems of majority instances and conflicts in overlapping regions in the dataset, reducing difficulties and distributing results in situations previously considered difficult, making it easy to identify the relevance of elements in the respective groups [29].

In Figures (2-5), the 2 LP Model presents sensitive behavior in relation to changes in the concentrations of elements in groups 1 and 2. Indicating growth in the categories MFC, MDC, MFE and MDE. Reducing the proportion in DC and indicating the growth of elements calculated by IRT as MFC and MFE.

This assignment of difficulty represented in the new quantities to be grouped by the clustering algorithms indicates that new elements were added in conflict regions and difficult to assign to the respective groups, but there was also a distribution of elements that were easier to assign, providing, when purchasing these assignments in the partitions, an evaluative way for human analysts to assign quality control to the properly grouped items, as well as the identification of optimal algorithms in solving the grouping objectives, providing a method to evaluate the decision-making of human analysts, in relation to the performance of the clustering algorithms. This action can be compared with situations that involve the construction of knowledge from reinforcement learning, as in the case of the impact of the use of differentiated methodologies in the educational environment and, in particular, the potential of gamification as an innovative strategy for teaching subjects such as mathematics [30].

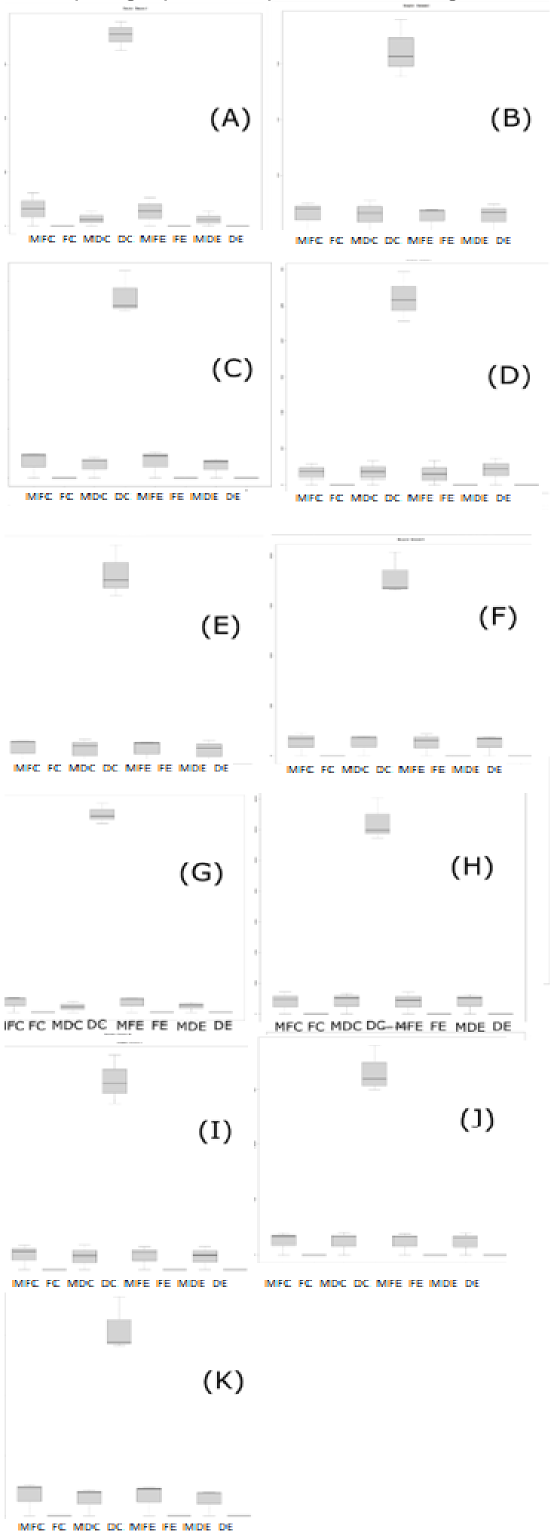
Figure 6 and 7 shows the boxplot graphs of the sums of the respective difficulty ranges obtained by the assignments of the clustering algorithms.

Figure 6 – Boxplot graphs of b-parameter categories in group 1.




Source: Author.

Figure 7 - Boxplot graphs of b-parameter categories in group 2.

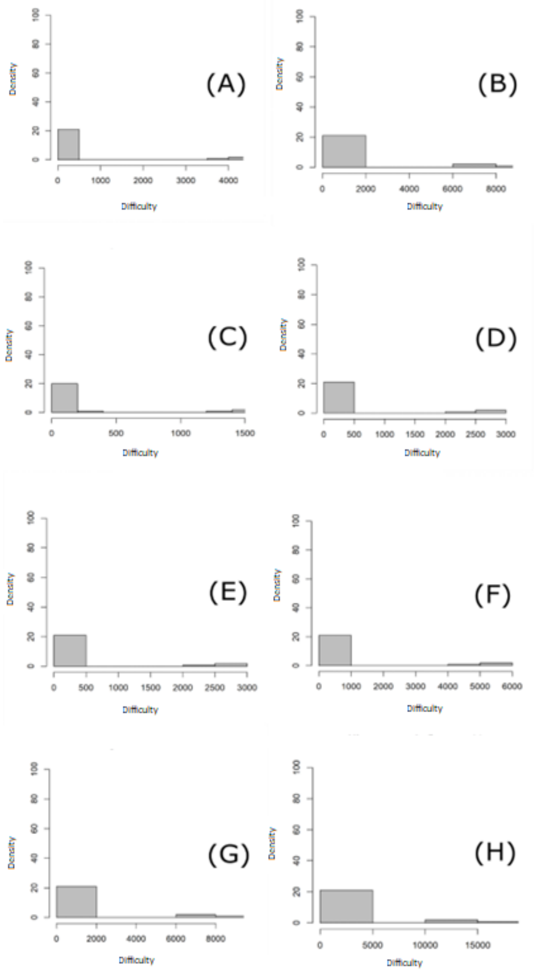



Source: Author.

It is also possible to observe that the FC, FE and DE distributions did not obtain records. This result can also be seen in the graphs in Figures 6 and Figure 7, thus representing an interval in which the data are recorded at the limits of the psychometric scale between Very Easy and Very Difficult to correctly or incorrectly assign the elements in the specific groups.

To understand these variations in concentrations, frequency graphs were generated, represented in Figure 8, where datasets 3, 4, 9 and 11 were compared by groups. Where it can be observed that the concentration of items is mostly at the moderate difficulty level (DC), recorded in Figures (5-8).

Figure 8 - Frequency graphs of b-parameter by dataset 3(A, C, E and G) – group 1 and dataset 3(B, D, F and H) – group 2, in datasets 3, 4, 9 e 11.




Source: Author.

Categorizing the classes, determined by the IRT Difficulty b-Parameter, into levels of complexity, makes it possible to identify that there is a gradation associated with the difficulty of the items to be grouped/classified, providing security in the application of the psychometric scale in classifications for Machine Learning algorithms for classification or group formation for clustering algorithms. Associating the computational learning process with the human one provides a cross-cultural educational experience, where there is a sensitive threshold: between never-ending data flows and algorithms that shape formative decisions, a new learning regime is insinuated in which meaning becomes more elusive than evident [31].

## CONCLUSION

The quality of the relevance is associated with the difficulties of these algorithms in correctly assigning elements to the clusters, and a psychometric scale based on IRT was used. This procedure was validated through comparisons of the metrics of statistical performance and internal and external validations of the clusters, evaluated in the data obtained from the tests of three IRT parameterization models: 1LP, 2LP and 3LP, focusing on the difficulty parameter, confirming the hypothesis that the IRT difficulty parameter can be successfully applied as a psychometric scale measure, to be used in the evaluation of clustering algorithms in complex environments to support the decision-making of human analysts. In addition, it was shown that it is possible to use IRT to determine optimal clustering algorithms in determining the relevance of the elements in the respective clusters.

## REFERENCES


1. J. Hathaliya and S. Tanwar, "An exhaustive survey on security and privacy issues in Healthcare 4.0". Computer Communications, vol. 153, n. September 2019, pp. 311–335, 2020, DOI: 10.1016/j.comcom.2020.02.018.

2. P. Wang, X. Yang, W. Ding, J. Zhan, and Y. Yao, “Three-way clustering: Foundations, survey and challenges”, Applied Soft Computing, vol. 151, n. December 2023, pp. 1-13, 2024a. DOI: 10.1016/j.asoc.2023.111131.

3. P. F. Leite-Filho, S. Melo, and R. Moura, “Métricas de rendimentos estatísticos em uma visão ampla dos dados”. In: Semana Universitária UPE 2021 - Democratizando a ciência do litoral ao sertão, 2021, Recife. Anais da Semana Universitária UPE 2021 - Democratizando a ciência do litoral ao sertão, pp. 1002, 2021b.

4. P. Leite-Filho and S. Melo. “Significance of Item Response Theory in Cluster Evaluation Metrics”. 2023 XLIX Latin American Computer Conference (CLEI), pp. 1-9, 2023, DOI: 10.1109/CLEI60451.2023.10346171.

5. A. Wartelle, F. Mourad-Chehade, F. Yalaoui, J. Chrusciel, D. Laplanche, and S. Sanchez. “Clustering of a health dataset using diagnosis co-occurrences”, Applied Sciences (Switzerland), vol. 11, n. 5, pp. 1–20, 2021. DOI: 10.3390/app11052373.

6. L. Ding, C. Li, D. Jin, and S. Ding, "Survey of spectral clustering based on graph theory." Pattern Recognition, vol. 151, n. January, pp. 1-12, 2024. DOI: 10.1016/j.patcog.2024.110366.

7. D. Gomes, J. Santos, J. Borges, M. Alves, and D. Andrade, A. Erdmann, "Theory of the Response To the Item in Research in Public Health". Journal of Nursing UFPE / Revista de Enfermagem UFPE, vol. 12, n. 6, pp. 1800–1812, 2018.

8. F. Martínez-Plumed, R. Prudêncio, A. Martínez-Usó, and J. Hernández-Orallo. “Item response theory in AI: Analysing machine learning classifiers at the instance level”. Artificial Intelligence, vol. 271, pp. 18–42, 2019. DOI: 10.1016/j.artint.2018.09.004.

9. Y. Chen, T. Silva Filho, R. Prudencio, T. Diethe and P. Flach."β3-IRT: A new item response model and its applications", Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS) 2019, Naha, Okinawa, Japan. PMLR: Volume 89.

10. Z. Chen and H. Ahn “Item Response Theory Based Ensemble in Machine Learning”. International Journal of Automation and Computing. 2020. Vol. 17. pp. 621-636.

11. S. Kandanaarachchi, "Unsupervised anomaly detection ensembles using item response theory", Information Sciences, vol. 587, pp. 142–163, 2022.

12. F. Baker and S. H. Kim, "The Basics of Item Response Theory”, Statistics for Social and Behavioral Sciences. 2nd ed. Cham, Switzerland, 2017.

13. J. Wu, J. Chen, H. Xiong and M. Xie, "External validation measures for K-means clustering: A data distribution perspective", Expert Systems with Applications, vol. 36, n. 3 PART 2, pp. 6050–6061, 2009. DOI: 10.1016/j.eswa.2008.06.093.

14. P. Jain, G. Shrivastava and U. Pandey, “Faculty Performance and Clusteirng – A Critical Review”. Journal od Algebraic Statistics, vol. 13, n. 1, pp. 111–120, 2022.

15. R. P. Chalmers. “Mirt: A multidimensional item response theory package for the R environment”. Journal of Statistical Software, v. 48, n. 6, pp. 1–29, 2012. DOI: 10.18637/jss.v048.i06.

16. S. Anand and S. Kumar, “Experimental Comparisons of Clustering Approaches for Data Representation”, ACM Computing Surveys, vol. 55, n. 3, pp. 1–33, 2023. DOI: 10.1145/3490384.

17. A. Feizollah, N. Anuar, R. Salleh, and F. Amalina, "Comparative study of k-means and mini batch k-means clustering algorithms in android malware detection using network traffic analysis", Proceedings - 2014 International Symposium on Biometrics and Security Technologies, ISBAST 2014, pp. 193–197, 2015. DOI: 10.1109/ISBAST.2014.7013120.

18. Y. Alomari, S. Abdullah, R. Zin, and K. Omar, “Iterative randomized irregular circular algorithm for proliferation rate estimation in brain tumor Ki-67 histology images”, Expert Systems with Applications, vol. 48, pp. 111–129, 2016. DOI: 10.1016/j.eswa.2015.11.012.

19. D. Sisodia, S. Verma, and O. Vyas, “A subtractive relational fuzzy c-medoids clustering approach to cluster web user sessions from web server logs”, International Journal of Applied Engineering Research, vol. 12, n. 7, pp. 1142–1150, 2017.

20. B. Desgraupes, "Clustering Indices". CRAN Package, vol. 1, n. April, pp. 1–34, 2013.

21. F. Wang, H. H, franco-penya, J. D. Kelleher, J. Pugh, and R, Ross. “An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity”. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), v. 10358 LNAI, n. July, pp. 291–305, 2017. DOI: 10.1007/978-3-319-62416-7_21.

22. P. Jain, G. Shrivastava and U. Pandey, "Faculty Performance and Clusteirng – A Critical Review". Journal od Algebraic Statistics, vol. 13, n. 1, pp. 111–120, 2022.

23. J. Thomas, M. Peñas and M. Mora, “New Version of Davies-Bouldin Index for Clustering Validation Based on Cylindrical Distance”, Proceedings - International Conference of the Chilean Computer Science Society, SCCC, n. 1, pp. 49–53, 2013. DOI: 10.1109/SCCC.2013.29.

24. N. Saini, D. Bansal, S. Saha and P. Bhattacharyya, “Multi-objective multi-view based search result clustering using differential evolution framework”, Expert Systems with Applications, vol. 168, n. November 2020, pp. 1-14, 2021. DOI: 10.1016/j.eswa.2020.114299.

25. R. Brouwer, “Extending the rand, adjusted rand and jaccard indices to fuzzy partitions”, Journal of Intelligent Information Systems, vol. 32, n. 3, pp. 213–235, 2009. DOI: 10.1007/s10844-008-0054-7.

26. D. Andrade and C. Valle, “Teoria da Resposta ao Item - Conceitos e Aplicações”, 1 ed, UNICAMP, 2000.

27. K. Wauters, P. Desmet, and W. V. D. Noortgate, “Item difficulty estimation: An auspicious collaboration between data and judgment”, Computers and Education, vol. 58, n. 4, pp. 1183–1193, 2012. DOI: 10.1016/j.compedu.2011.11.020.

28. P. Leite-Filho and S. Melo, “Tomada de decisões baseada no parâmetro b da teoria da resposta ao item nas classificações de ensemble”, Revista dos Mestrados Profissionais, vol. 11, n. 2, pp. 235–249, 2022. DOI: 10.51359/2317-0115.2022.256869.

29. A. Chakraborty, K. Ghosh; R. De, E. Cuevas, and R. SARKAR, “Learning automata based particle swarm optimization for solving class imbalance problem”, Applied Soft Computing, vol. 113, pp. 1–9, 2021. DOI: 10.1016/j.asoc.2021.107959.

30. L. L. Rodrigues; S. DE Oliveira; E, S, Soares; R. I. DA Silva; R. S. Fortes. “Gamification applied to the teaching of first-degree equations using a virtual torque balance”. ARACÊ , , v. 7, n. 6, pp. 29138–29174, 2025. DOI: 10.56238/arev7n6-007.

31. C. R. Frias; N. B. DOS Reis; L. S. Azevedo; G. M. F. DE Souza; E. M. DE Matos; D. P. Ramos; K. O. N. Carvalho; L. R. DA Silva. “Education and technology: challenges and possibilities the educational experience in the age of digital platforms”. ARACÊ , v. 7, n. 6, pp. 29191–29207, 2025. DOI: 10.56238/arev7n6-009.