Implementation of Clustering Techniques to Data Obtained from a Memory Match Game Oriented to the Cognitive Function of Attention

Marcos Orellana(✉), María-Inés Acosta-Urigüen, and Reinerio Rodríguez García

Universidad del Azuay, Cuenca, Ecuador
{marore,macosta}@uazuay.edu.ec, rrodriguez@es.uazuay.edu.ec

Abstract. Serious games are software applications with an explicit educational objective that have been thoroughly thought out and designed as a learning instrument or tool. They allow the user to experience situations similar to real life and learn from their mistakes through immediate feedback. These games have been developed in various fields, including business, industry, marketing, health, government, among others. For instance, some are used for cognitive training in human beings, where attention and memory are fundamental axes during the various stages of the human life cycle. In this context, the "memory match game" is a card game where many pairs of cards are laid face down, and its objective is to match pairs in the lowest time and with the minimum number of wrong clicks. This information is registered and stored along with the player's sociodemographic information. Thus, this article aims to analyze the dataset, applying clustering techniques in order to find behavioral patterns. The age variable was used to generate 4 age groups that served as the basis for applying the unsupervised machine learning algorithm, k-means. The results show the behavior of the data in relation to the age groups, it is evident that the more experience the players gain, times and scores improve, regardless of age.

Keywords: Serious games · K-means · Memory match game

1 Introduction

Cognitive functions directly influence educational and professional performance, socioeconomic achievement, health, and longevity. The study of the decrease in these capacities, although related to age, is also associated with deficiencies in the performance of daily activities when these fall outside normal ranges. In general, mild deficiencies in the aforementioned functions are considered normal in aging [1].

The cognitive functions include processing speed, attention, working memory, verbal and visual learning, and executive functions. Attention has components that measure vigilance, orientation, and detection. Detection is the component that quantifies the focus of attention on a stimulus for a time, even with the presence of distractions or

J. Herrera-Tapia et al. (Eds.): TICEC 2022, CCIS 1648, pp. 201–216, 2022.
https://doi.org/10.1007/978-3-031-18272-3_14

increased fatigue. For example, vigilance or sustained attention can be measured using sets of identical pairs, selective attention can be measured using the Stroop test [2, 3].

Serious Games (SG) are computer applications with purposes beyond the entertainment that can register and store data transparently during the gameplay [4, 5]. SG are mainly developed in the fields of military, health, science, and education. These games can include enhanced and interactive gameplay, levels and quests, and interaction with game objects [5, 6]. Some SG have been developed to support cognitive stimulation, training, and rehabilitation [6, 7], others try to modify the behavior or improve the health care [8].

Data science is an area of study focused on the prediction, exploration, and understanding of data [9]. The algorithms and techniques include supervised algorithms, unsupervised algorithms, and visualization techniques [10].

Unsupervised algorithms base the training process without previously defined labels or classes applied to the dataset. Clustering techniques and algorithms are considered unsupervised, and they measure the distance between two entities in a dataset, and, based on it, the clusters are formed [11]. K-means is an unsupervised algorithm which describes simple mathematical processes and has a fast convergence [12, 13].

The aim of this paper is to take, as an starting point, the results obtained by [14] and apply clustering techniques considering the level of the game in which the player is in order to evaluate the performance.

The paper is structured as follows: Sect. 2 presents the related works for the use of clustering algorithms in the field of serious games, Sect. 3 describes the methodology based on SPEM and its implementation, and finally, Sect. 4 presents the conclusions and future work.

2 Related Works

The implementation of data science techniques for the analysis of data from SG has been widely analyzed. For example, [15] presented a SG to predict math skill level of students with special capacities. The game consisted of a quiz with multiple choices (4 options) where each question had a predefined time limit. The quiz was applied to a sample of children with normal and special capacities, moreover, sociodemographic information was collected, along with the answers. To classify data, several ranges and percentages were calculated, and six data mining methods were used to analyze the data. The conclusions showed the correlation between sociodemographic variables when determining the level of math skills for the normal student, but this could not be found in the students with special needs group. The best accuracy was obtained using the JRIP algorithm.

The authors of [16] presented a SG to collect, analyze and visualize body movement information based on the data collected for the SG Hammer and Planks, whose objective is to rehabilitate patients with balance disorders. The user has to move the body in different directions in a 2D environment simulating the navigation on a ship. The hierarchical clustering algorithm was used to identify similar distribution of movements. However, the work does not mention scores related to the game performance, although it describes some visualization techniques to show the evolution of movements.

An adaptive rehabilitation bot based on the use of SG was proposed by [17]. RehaBot was developed for traditional therapy solutions. The application can adjust therapies to cover the whole body; includes a virtual assistant to address patients to perform the exercises correctly through 3D illustrations, and adjust the game's difficulty level in real-time according to the patients' abilities. The work presents the use of the mining techniques to predict improvement based on a schedule of exercises, but it does not describe the application of the techniques themselves.

A SG for science and technology presents a virtual world where players explore thematic islands to discover games, news, photos, and videos [18]. The players' information, performance, and interactions were stored and analyzed; the processes of detecting and removing outliers and extreme values were carried out; linear regression and clustering techniques were also applied. K-means technique was used, considering one dependent variable with two influential independent variables. The results demonstrate that the number of accesses to the game, the quests visited, and the advantages used are important factors that influence scores and time playing; the clustering allows to identify beginner, intermediate and advanced players according to their experience.

The authors of [14] presented a methodology to evaluate data from SG. The proposal included four stages, and the implemented data mining technique was clustering. Although several demographic variables were collected, only time, gender, and level were analyzed. The results showed a strong relationship between age and gender, and the proposed future work aims to compare models and techniques to obtain better results.

3 Application of the Methodology

The methodology is proposed in Fig. 1 through a diagram with the software process and the metamodel engineering systems 2.0 (SPEM 2.0). According to [19], the activities in the methodology describe the input and output of devices in the modeling process.

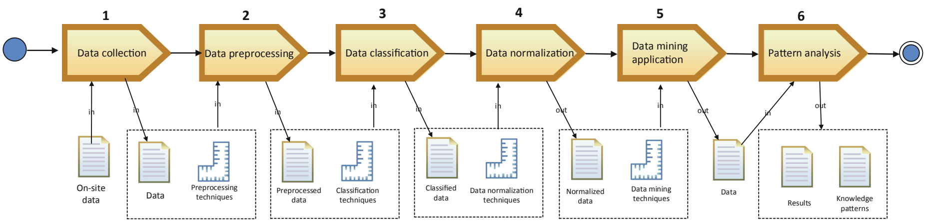


Fig. 1. Methodology to figure out knowledge patterns in serious games data.

The methodology is divided into data collection, processing, classification, normalization, mining techniques, and analysis. Data collection focuses on using in situ criteria for serious games data collection. The game of pairs consists of showing the tokens in a specific time and location for these to be covered for the subject to remember the initial position [20]. Sociodemographic variables and SG data execution variables are collected and stored in a single online data set for all players. The game is available in the website https://jserionew-8e818.web.app/#/.

The data processing stage describes a group of unprocessed data and its preliminary preparation by applying processing techniques such as the elimination of missing, faulty, extreme, among others, data values to obtain a clear data set. Within the collected data, there are several inconsistencies such as missing names, age, birth date or the time of a single game level. Lastly, the extreme values register time beyond half an hour to finish each level; these directly affect the participant's score. These registers are eliminated to obtain a clean data set and efficient results.

According to Hou, there is a data classification aimed to categorize the results because people's attention changes according to age [21]. Also, [22] applies a data classification based on the results of the participants' scores to group them in relation to their performance. For this study, four ranges have been considered: kids (up to 11 years), teenagers (from 12 to 20 years), adults (from 21 to 59 years), and elderly (from and beyond 60).

Data normalization applies the RapidMiner program to transform the scores and age in a standard scale for these two attributes using a value range from 0 to 1 [23].

This paper uses K-clustering and Neural Networks as data mining techniques to analyze and verify the most relevant technique to find efficient behaviors. This section uses K-clustering available in RapidMiner [24] to determine the K number of clusters to satisfy the criteria [25]; the K algorithm is a grouping method that allows to work in large data sets with significant values because it is capable to provide an efficient classification [21].

The data analysis stage uses the results from the data mining techniques. The data is statistically analyzed to determine the best distribution. In Fig. 2, it is possible to observe the quantity of score values obtained by the participants whose most frequent score is 1800 points; this is a significant value considering that the minimum score is 1000 and the maximum is 2000. The mean score of all participants is 1704.57, which is an acceptable value as the perfect score is 2000. The standard deviation is 208.62, so it is a relevant value to analyze and quantify in the data dispersion.

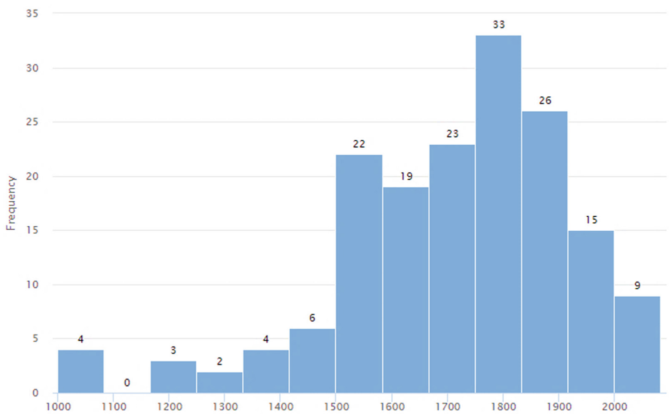


Fig. 2. Frequency of scores for all participants.

Table 1 presents the data distribution, where most data come from the adult category (21–58 years) with 67.66% (n = 112) of participants. For the other categories, teenagers (12–20 years) represent 19.16% (n = 32), kids represent 10.78% (n = 18), and elderly (beyond 60 years) represent 2.40% (n = 4).

Table 1. Data distribution of participants according to categories.

Category	Quantity	Percentage
Kids (up to 11 years)	18	10.78%
Teenagers (12–20 years)	32	19.16%
Adults (21–59 years)	112	67.66%
Elderly (from and beyond 60 years)	4	2.40%

Figure 3 presents a box plot of the categorized ages and scores to interpret the scores, laying out data quartiles and atypical values.

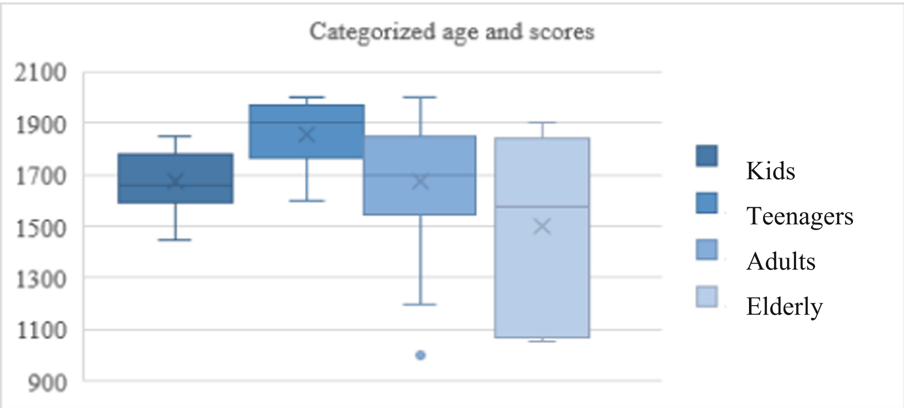


Fig. 3. Box plot of categories and scores.

The box plot portraying the Kids category is compact and not dispersed; the same pattern repeats for Teenagers and Adults. Conversely, in the Elderly category, there is more data dispersion due to the diversity in their scores, considering the lower quartile score and the minimum score compared to the other categories. This is due to the lower number of samples in the category, so a lower score affects all the results. Conversely, the Teenagers category has higher scores and less data dispersion; thus, this category's results show better attention in teenagers, representing a greater performance than the other categories.

Figure 4 presents the trials made by participants in the Kids category (0–11 years) in which five people made three trials, two people made four trials, and one person madefive5 trials in each of the 4 phases of the serious game. Every phase, according to the number of trials, shows an improvement in the time that take to finish each game's phase. As a result, they improve their attention. In the first phase of the game there is a slight increase in the time, but this increase is vastly irrelevant.

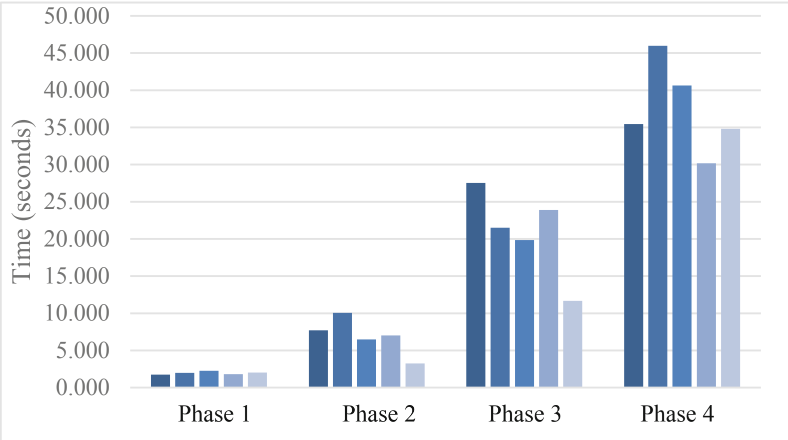


Fig. 4. Trials by the participants in the kids category.

Figure 5 depicts the trials made by the Teenagers category (12–20), in which 30 people made one trial and 2 made two trials of each of the 4 phases of the serious game. It can be appreciated that the time to finish each game's stages improves in each phase and according to the trial number.

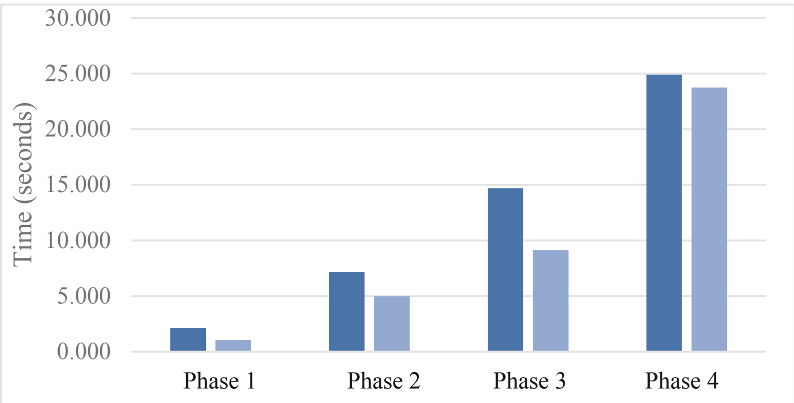


Fig. 5. Trials by the participants in the teenagers category.

In the Adults category, Fig. 6 portrays the trials performed by its participants in which 73 people made one trial, 15 people made two trials, seven people made three trials, and one person made 11 trials for each one of the phases of the serious game. In each phase, the time taken to finish the serious game improved in relation to the number of trials.

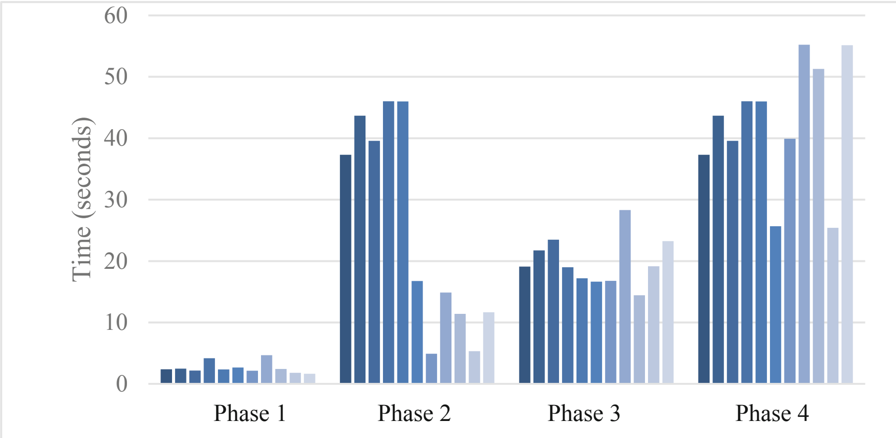


Fig. 6. Trials by the participants in the adults category.

In the category of Elderly (beyond 60 years), Fig. 7 shows the trials performed by its participants in which three people made one trial, and two made two trials in each of the SG phases. It is noticed that the time to finish the game improves as the number of trials increases.

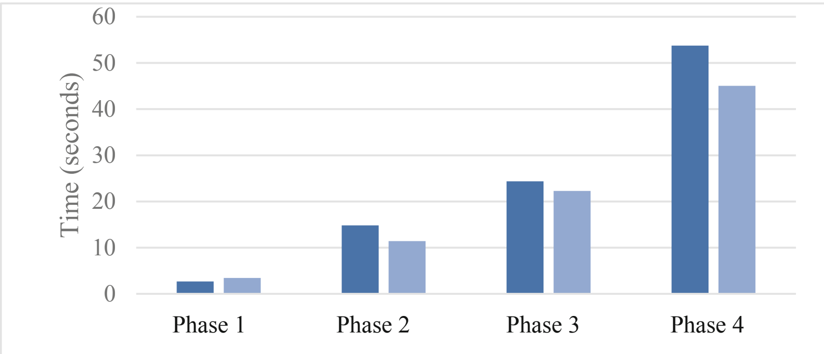
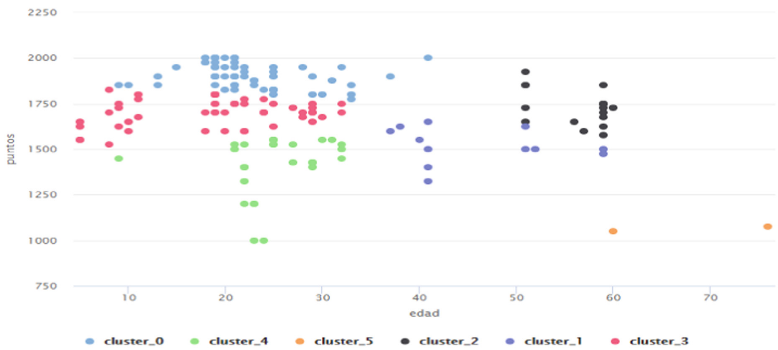


Fig. 7. Trials by the participants in the elderly category.

3.1 Cluster Application

For data classification, the Euclidean distance is used as a measurement of association. The formed groups contain similar individuals, so the distance between them is short.

Figure 8 shows that when the group of ages and participants' scores is classified, these can be divided into groups with similar values; cluster 5 is separated due to its two atypical values. A large amount of the values is concentrated within the range of less than 41 years and more than 1350 points. As a result, the vast majority of people younger than 40 years have higher scores, and after this age, other factors may affect the performance and efficiency to finish the game.



Cluster	Quantity	Percentage
0	57	34.3%
1	12	7.20%
2	17	10.20%
3	51	30.70%
5	27	16.30%
6	2	1.20%

Fig. 8. Cluster of age and score of the participants.

Figure 9 portrays the classification of data according to age and time of the first phase of the game, where the participants are divided into two groups defined by score similarity, defined according to their age and time into four groups. Also, cluster 2 is differentiated by an atypical value. A great quantity of the values concentrates in cluster zero with 126 data points, equivalent to 76% of the data set, meaning that people younger than 40 years end the first game's phase within 4 s.

Figure 10 displays the data set classification of age and time of the second phase of the game, which divides the participants into groups defined by the similarity of the values, thus, identifying four groups. Also, cluster 2 is differentiated due to an atypical value. A significant number of the values concentrate in cluster zero with 120 data points, equivalent to 72.3% of the data set, meaning that people younger than 37 years finish the second phase of the game in less than 14 s. Subsequently, there are dispersed values in cluster 1 for people between the ages of 20 and 26 years who exceed the 14 s.

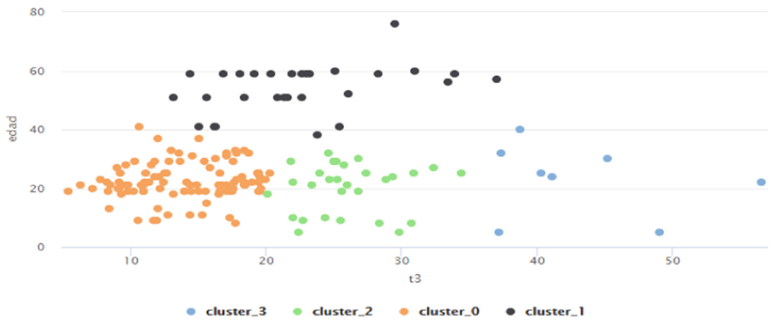
Cluster	Quantity	Percentage
0	126	76%
1	26	16%
2	1	1%
3	13	8%

Fig. 9. Data classification by age and time of the first phase.

Cluster	Quantity	Percentage
0	120	72.3%
1	20	12%
2	1	0.6%
3	25	15.1%

Fig. 10. Data classification by age and time of the second phase.

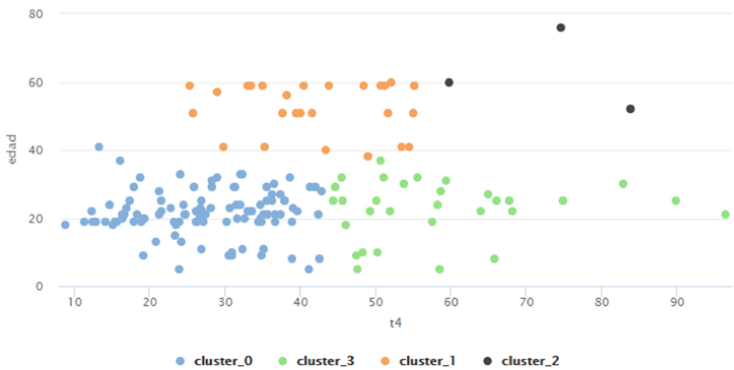
Figure 11 depicts the data set classification of age and time of the third phase of the game, which divides the participants into groups defined by the similarity of the values; thus, identifying four cluster groups with dispersion. The largest concentration of data points is in cluster zero, with 59.6% of the data set, meaning that people older than 40 years finish the third phase of the game in 22 s or less. Then, cluster 1 is very close since it represents people older than 40 years with an estimated time between 12 and 38 s. Moreover, cluster 2 involves people between 5 and 38 years with a time between 20 and 35 s. Finally, cluster 3 has 8% of the data set depicting people between 5 and 40 years that take between 35 and 60 s to finish the game.



Cluster	Quantity	Percentage
0	99	59.6%
1	29	17.5%
2	30	18.1%
3	8	4.8%

Fig. 11. Data classification by age and time of the third phase.

Figure 12 depicts the data set classification of age and time of the third phase of the game, which divides the participants into groups defined by the similarity of the values; thus, identifying three cluster groups with dispersion. The largest concentration of data points is in cluster zero with 63.3% of the data set, meaning that people until 41 years finish the fourth phase of the game in less than 45 s. Then, cluster 1 represents 16.3% of the data set with people between 39 and 60 years with an estimated time between 25 and 45 s. Moreover, cluster 2 represents 1.8% of people between 55 and 68 years with a time between 60 and 85 s. Finally, cluster 3 has 18.7% of the data set depicting people between 5 and 39 years that take between 45 and 95 s to finish the game's phase.



Cluster	Quantity	Percentage
0	105	63.3%
1	27	16.3%
2	3	1.8%
3	31	18.7%

Fig. 12. Data classification by age and time of the fourth phase.

To obtain data in the same scale, a normalization technique was applied to standardize the values in a common scale for the time variables for each game level (TPN1, TPN2, TPN3, TPN4), ages, scores, and sex. The objective is to analyze the data efficiently.

Once the data is normalized, a clustering technique is applied to classify it into four groups. Table 2 portrays the centroids of the four clusters and the different game variables.

Table 2. Normalized and classified data centroids.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Trial	0.110	0.150	0.052	0.037
Age	0.311	0.516	0.203	0.289
Score	0.536	0.551	0.809	0.839
TPN1	0.201	0.243	0.193	0.132
TPN2	0.156	0.189	0.102	0.108
TPN3	0.424	0.493	0.198	0.162
TPN4	0.416	0.493	0.210	0.221
Sex	0	1	1	0

Figure 13 shows that the time variables in each level of the game, age, and sex are important in each cluster classification. This is why the analysis is more thorough in the trials of the participants in each level of the game. The aim is to verify the people in each category.

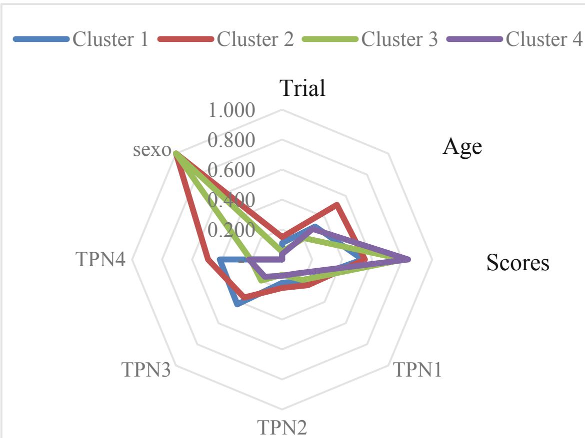
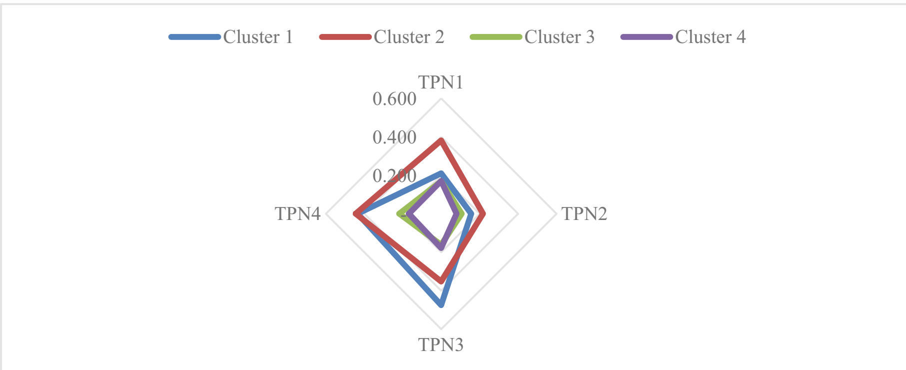


Fig. 13. Radial graph of the application of clustering with data from the centroids of the demographic variables.

Figures 14, 15, and 16 have centroids of the groups that include time variables in the different game levels (TPN1, TPN2, TPN3, TPN4) of each one of the trials. They show graphs that represent the classification of the data in each one of the trials of the game. From these, cluster 1 has a significant influence over the others, so it is thoroughly analyzed to identify the category of the people's ages in this classification. Tables 17, 19, and 21 reflect the different categories of age, quantity, and percentage of men and women found, the mean time of men (MTM) and the mean time of women (MTW) determine the time taken to finish the different levels of the game. Lastly, the mean scores of men (MSM) and mean scores of women (MSW) are portrayed too (Table 3).

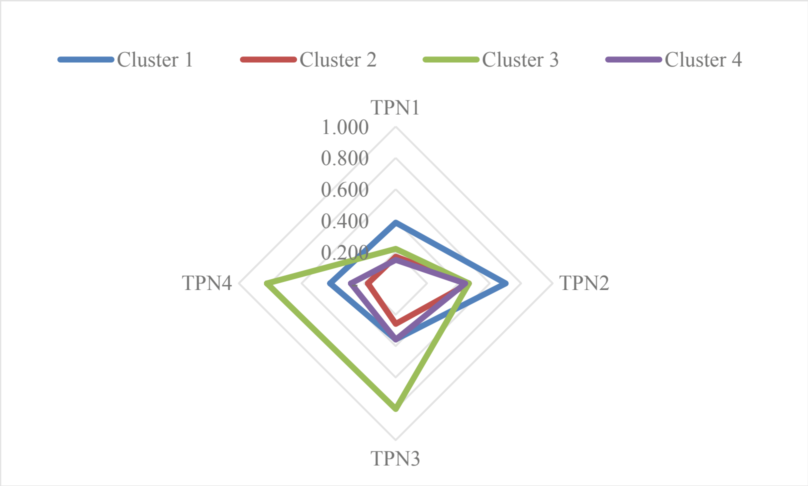
Table 3. Results of the data for categories in Figure 15.

Category	Data quantity	Men quantity	Women quantity	MTM	MTW	MSM	MSW
Kids	5	2	3	16.4	19.2	1612.5	1725
Teenagers	30	22	8	12.0	12.9	1843.2	1890
Adults	73	35	38	17.8	15.6	1649.3	1718
Elderly	2	2	0	29.8		1062.6	
Total	110	61	49				
Mean				18.99	15.90	1541.9	1778



	Cluster 1	Cluster 2	Cluster 3	Cluster 4
TPN1	0.201	0.381	0.169	0.171
TPN2	0.156	0.217	0.107	0.080
TPN3	0.475	0.352	0.161	0.179
TPN4	0.435	0.445	0.218	0.168

Fig. 14. Graph of grouped centroids in trial 1.

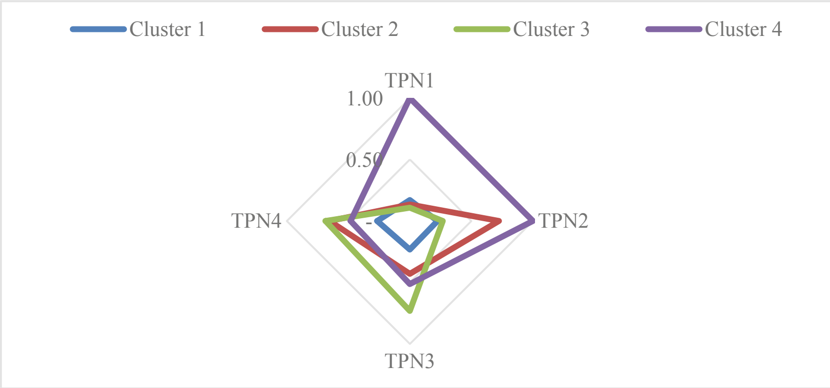


	Cluster 1	Cluster 2	Cluster 3	Cluster 4
TPN1	0.387	0.170	0.220	0.149
TPN2	0.701	0.466	0.467	0.443
TPN3	0.357	0.259	0.801	0.359
TPN4	0.418	0.178	0.821	0.285

Fig. 15. Graph of grouped centroids in trial 2.

Table 4. Results of the data for categories of trial 2 (Figure 15).

Category	Data quantity	Men quantity	Women quantity	MTM	MTW	MSM	MSW
Kids	5	3	2	19.77	20.062	1608.3	1650
Teenagers	2	2	1	9.59	9.862	1850	1950
Adults	16	10	6	19.94	16.928	1725	1579.2
Elderly	1	1	0	23.34		1725	
Total	24	15	9				
Mean				18.2	15.6	1727.08	1726.40



	Cluster 1	Cluster 2	Cluster 3	Cluster 4
TPN1	0.17	0.13	0.11	1.00
TPN2	0.22	0.72	0.27	1.00
TPN3	0.23	0.43	0.73	0.51
TPN4	0.27	0.65	0.69	0.48

Fig. 16. Graph of grouped centroids in trial 3

Table 5. Results of the data for categories of trial 2 (Figure 15).

Category	Data quantity	Men quantity	Women quantity	MTM	MTW	MSM	MSW
Kids	5	3	2	13.81	22.54	1783.3	1575
Adults	8	6	2	18.67	22.74	1716.6	1475
Total	13	9	4				
Mean				16.2	22.6	1750.5	1525

4 Conclusion

The data obtained from the serious game to assess attention was analyzed using data mining to understand the relevance within the variables of age, sex, score, and time to finish the game's levels. The clustering graphs draw two distinct patterns for men and women in relation to their mean time MTM and MTW and mean score MSM and MSW. While men finish the game faster than women, their score is lower than women. In Table 4, the data of the second trial, the mean´s difference in the sex category is irrelevant even if men take less time to finish than women but obtain a lower score. This data sample shows that adult men take, in average more time to finish than adult women. Also, Table 5 shows how men's time is lower and their scores higher in the Kids and Adults categories compared to women.

The data evaluation, through its grouping and normalization, determines that participants become more experienced as they play the game more, which improves the time and score. The age category is the most influential for the results because kids, teenagers, adults, and the elderly have a different attention span.

The future works aim to apply other techniques besides clustering, incorporating the analysis of other demographic variables such as level of education or health condition, and to increase the levels of difficulty that the SG has. Variables such as body mass index, exercise time per week and educational or work level will be analyzed to identify their importance in relation to the results obtained in serious play itself.

References

1. Lövdén, M., Fratiglioni, L., Glymour, M.M., Lindenberger, U., Tucker-Drob, E.M.: Education and cognitive functioning across the life span. Psychol. Sci. Public Interes. **21**(1), 6–41 (2020). https://doi.org/10.1177/1529100620920576
2. Burdick, K.E., Russo, M., Martin, J.: Neuropsychological assessment and psychological tests. Mt. Sinai Expert Guid., 24–27 (2016). https://doi.org/10.1002/9781118654231.ch5
3. Esterman, M., Rothlein, D.: Models of sustained attention. Curr. Opin. Psychol. **29**, 174–180 (2019). https://doi.org/10.1016/j.copsyc.2019.03.005
4. Alonso-Fernández, C., Martínez-Ortiz, I., Caballero, R., Freire, M., Fernández-Manjón, B.: Predicting students' knowledge after playing a serious game based on learning analytics data: a case study. J. Comput. Assist. Learn. **36**(3), 350–358 (2020). https://doi.org/10.1111/jcal.12405
5. Vidakis, N., Barianos, A.K., Trampas, A.M., Papadakis, S., Kalogiannakis, M., Vassilakis, K.: In-game raw data collection and visualization in the context of the "ThimelEdu" educational game. In: Lane, H.C., Zvacek, S., Uhomoibhi, J. (eds.) CSEDU 2019. CCIS, vol. 1220, pp. 629–646. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58459-7_30
6. Rego, P., Moreira, P.M., Reis, L.P.: Serious games for rehabilitation a survey and a classification towards a taxonomy (2002)
7. Palumbo, V., Paternò, F.: Serious games to cognitively stimulate older adults: a systematic literature review. In: Proceedings of the ACM International Conference on Proceeding Series, pp. 199–208 (2020). https://doi.org/10.1145/3389189.3393739
8. Yáñez-Gómez, R., Cascado-Caballero, D., Sevillano, J.-L.: Academic methods for usability evaluation of serious games: a systematic review. Multimedia Tools Appl. **76**(4), 5755–5784 (2016). https://doi.org/10.1007/s11042-016-3845-9
9. Blei, D.M., Smyth, P.: Science and data science. Proc. Natl. Acad. Sci. U. S. A. **114**(33), 8689–8692 (2017). https://doi.org/10.1073/pnas.1702076114
10. Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., Fernández-Manjón, B.: Applications of data science to game learning analytics data: a systematic literature review. Comput. Educ. **141**, 103612 (2019). https://doi.org/10.1016/j.compedu.2019.103612
11. Kandasamy, I., Kandasamy, W.B.V., Obbineni, J.M., Smarandache, F.: Indeterminate likert scale: feedback based on neutrosophy, its distance measures and clustering algorithm. Soft. Comput. **24**(10), 7459–7468 (2019). https://doi.org/10.1007/s00500-019-04372-x
12. Yuan, C., Yang, H.: Research on K-value selection method of K-means clustering algorithm. J. **2**(2), 226–235 (2019). https://doi.org/10.3390/j2020016
13. Li, Y., Wu, H.: A clustering method based on K-means algorithm. Phys. Procedia **25**, 1104–1109 (2012). https://doi.org/10.1016/j.phpro.2012.03.206
14. Orellana, M., Lima, J.-F., Acosta Urigüen, M.-I., Patiño, A., Álvarez, N., Cordero, J.: Data mining applied to a serious game of memory and attention training. In: Narváez, F.R., Proaño, J., Morillo, P., Vallejo, D., González Montoya, D., Díaz, G.M. (eds.) SmartTech-IC 2021. CCIS, vol. 1532, pp. 58–68. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-99170-8_5

15. Yuhana, U.L., Mangowal, R.G., Rochimah, S., Yuniarno, E.M., Purnomo, M.H.: Predicting Math performance of children with special needs based on serious game (2017). https://doi.org/10.1109/SeGAH.2017.7939276
16. Purwantiningsih, O., Sallaberry, A., Andary, S., Seilles, A., Azé, J.: Visual analysis of body movement in serious games for healthcare. In: Proceedings of the IEEE Pacific Visualization Symposium, pp. 229–233 (2016). https://doi.org/10.1109/PACIFICVIS.2016.7465276
17. Afyouni, I., Murad, A., Einea, A.: Adaptive rehabilitation bots in serious games. Sensors (Switzerland) **20**(24), 1–30 (2020). https://doi.org/10.3390/s20247037
18. Benmakrelouf, S., Mezghani, N., Kara, N.: Towards the identification of players' profiles using game's data analysis based on regression model and clustering. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015, pp. 1403–1410 (2015). https://doi.org/10.1145/2808797.2809429
19. Standars Development Organization: Software & Systems Process Engineering Metamodel. https://www.omg.org/spec/SPEM/2.0/About-SPEM/
20. Khenissi, M.A., Essalmi, F., Jemni, M.: A learning version of memory match game. In: Proceedings of the IEEE 14th International Conference on Advanced Learning Technologies ICALT 2014, pp. 209–210 (2014). https://doi.org/10.1109/ICALT.2014.67
21. Hou, H.T.: Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: a video-based process exploration. Comput. Human Behav. **48**, 424–435 (2015). https://doi.org/10.1016/j.chb.2015.02.010
22. Slimani, A., Elouaai, F., Elaachak, L., Yedri, O.B., Bouhorma, M.: Learning analytics through serious games: data mining algorithms for performance measurement and improvement purposes. Int. J. Emerg. Technol. Learn. **13**(1), 46–64 (2018). https://doi.org/10.3991/ijet.v13i01.7518
23. RapidMiner Named a Leader in The Forrester Wave^{TM}: Multimodal Predictive Analytics And Machine Learning, Q3 2020. RapidMiner (2020). https://rapidminer.com/news/rapidminer-named-leader-multimodal-predictive-analytics-machine-learning/
24. RapidMiner Inc.: RapidMiner _ Best Data Science & Machine Learning Platform, 2020 (2020)
25. Li, Y.G.: A clustering method based on K-means algorithm. Appl. Mech. Mater. **380–384**, 1697–1700 (2013). https://doi.org/10.4028/www.scientific.net/AMM.380-384.1697