K-Means and Fuzzy C-Means Optimization using Genetic Algorithm for Clustering Questions

Article · December 2021

DOI: 10.47679/ijasca.v1i1.2

Agung Bella Putra Utama
State University of Malang

Prasetya Widiharso
State University of Malang

International Journal of Advanced Science and Computer Applications (2022),1(1): 1-10
DOI: 10.47679/ijasca.v1i1.2
https://ijasca.org/index.php/ijasca

RESEARCH ARTICLE

K-Means and Fuzzy C-Means Optimization using Genetic Algorithm for Clustering Questions

Siti Sendari[a,1*], Agung Bella Putra Utama [a,2], Nastiti Susetyo Fanany Putri [a,3], Prasetya Widiharso [a,4], Rizki Jumadil Putra [a,5]

Published online: 10 January 2022

Abstract

The grouping of data can be used in the development strategy of an educational game application. The process of grouping data that initially behaved differently into several groups that now behaved more uniformly. As well as grouping the data on the difficulty level of the questions on the educational game question board. This grouping of questions is needed to get the dominant values that will be the characteristics of each group of questions that exist. The clustering method is quite widely used to overcome problems related to data grouping. This clustering is a method of grouping based on the size of the proximity, the more accurate the cluster formed, the clearer the similarity of the difficulty level of the questions. Thus, educational game developers can determine the strategy for placing the existing questions more precisely. Many clustering methods can be used to group the data on this question, including K-Means and Fuzzy C-Means (FCM) which are then optimized using the Algorithm Genetics. From the results of the research conducted, optimization gives better results for clustering questions.

Keyword: Educational Game; Clustering; Data Mining; Fuzzy C-Means; Genetic Algorithm

Introduction

Games are one of the most popular entertainment media today [1]. Games can be played by anyone, both children and adults, depending on the type of game chosen [2]. One type of game is an educational game. Educational games are classified as innovations in the world of education, which are a combination of learning and playing [3]. Today's technological developments also affect the development of educational games, where what was once only a conventional game is now developing into a multimedia-based game. Examples of educational games include RPM games, puzzle games, mix and match, and so on.

The discussion of educational games, will not be separated from the discussion about the questions inserted in it [4]. This question is usually used as a benchmark for players' understanding of the material in the game. Currently, there is no definite way to arrange evaluation questions in educational games [5]. For this reason, this study will try to apply the fuzzy C-Means clustering algorithm to find clusters of the difficulty level of the items to be inserted. With this cluster, game developers will find it easier to distribute questions in the games they make. In this case, the development of the LantaSmart game was carried out by adopting a display in the form of a snake and ladder. Where, the questions will be clustered into three levels, namely easy, medium, and difficult. Questions with easy clusters will be placed in squares with snake tails, questions with medium clusters will be placed in general plots, and questions with difficult clusters will be placed in squares containing ladders which are also used as a reward for giving correct answers. So that the existing clustering process can maximize the placement of questions in the LantaSmart game.

The purpose of the research is to apply clustering algorithms, especially K-means and fuzzy C-means in clustering evaluation questions in the development of the LantaSmart educational game. K-Means is sensitive to initialization at the beginning of the cluster center and is more optimal in a local cluster [6]. Meanwhile, fuzzy C-means have advantages in clustering with more than one variable and depend on the degree of membership. So that a very good genetic algorithm is applied to help find the initial cluster center on K-means and determine the degree of initial membership [7]. The K-Means genetic algorithm can produce groupings with a better level of clustering variation than the simple K-Means algorithm as well as the fuzzy C-means algorithm [8]. In this study, clustering was carried out using k-means and optimization of

[1-5] Faculty of Engineering, Universitas Negeri Malang

*) corresponding author

Siti Sendari

Faculty of Engineering, Universitas Negeri Malang, Malang, Indonesia

Email: siti.sendari.ft@um.ac.id

centroid determination in the k-means process using a genetic algorithm and as a determinant of the initial membership degree of fuzzy C-means.

Methods

The method proposed in the clustering research about the LantaSmart educational game application is the K-Means and Fuzzy C-Means (FCM) algorithms which are optimized with the Genetic Algorithm.

A. LantaSmart Games

LantaSmart game is an educational game that contains material and practice questions about traffic signs. The LantaSmart game is also made as a learning medium for elementary school children so that they can recognize and know the rules of traffic signs from an early age [9], [10], to reduce the number of violators and traffic accidents that exist [11]–[13]. The questions in the LantaSmart game are used as a benchmark for user understanding of the material that has been studied. The questions are packaged in the form of a snake and ladder game to increase interest and enthusiasm for learning for users as shown in Figure 1.

Fig. 1. The appearance of the snake and ladder in the LantaSmart game

The placement of the questions in each of the plots on the snake and ladder will be adjusted to the results of clustering the level of difficulty of the questions which will be divided into three parts, namely easy, medium, and difficult. For questions with easy clusters, they will be placed in plots with snake tails, so that if users cannot answer questions in easy clusters, they will get a penalty for going down to the bottom square. Questions with medium clusters will be placed in general plots. For questions with difficult clusters, they will be placed in a plot containing stairs in it, it aims to give a reward as an achievement obtained by the user for giving the correct answer.

B. Data Research

Table 1. Dataset

Question Number	Right Answer	Wrong Answer
1	26	7
2	28	5
3	33	0
4	23	10
5	8	25
…	…	…
…	…	…
46	23	10
47	27	6
48	23	10
49	20	13
50	21	12

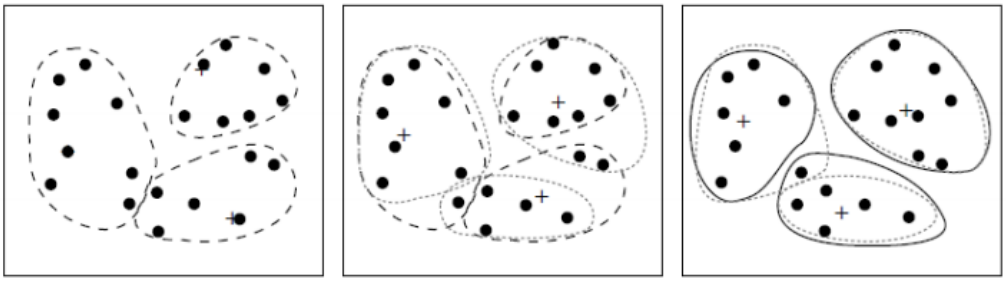
The data used in this study is data from the corrected answers to questions that have been done by 33 elementary school students consisting of male or female students with a range of ages from 8-12 years. The questions have given consist of 50 numbers in a multiple-choice format containing questions about traffic signs. The method used to obtain this data is by distributing questionnaires to the existing draft questions directly to students. From the results of the correction of the existing answers, they are then processed into a dataset so that they can be used for the clustering process of the difficulty level of the existing questions as can be seen in Table 1.

C. Question Clustering Method

The clustering method is quite widely used to overcome problems related to data segmentation. This clustering is a grouping method based on the size of proximity (similarity) where the clusters do not have to be the same but are grouped based on the proximity of an existing sample data characteristic, one of which is by using the euclidean distance formula [14]. The following is a clustering method that can be used to group questions in the LantaSmart educational game application, including:

1. *K-Means*

K-means is the most common and simple clustering technique. The purpose of this clustering is to group objects into k clusters or groups as shown in Figure 2



(a) Initial clustering (b) Iterate (c) Final clustering

Fig. 2. K Means Clustering

2. *Fuzzy C-Means*

The Fuzzy C-Means algorithm is a data clustering technique where the determination of the point of each data in a cluster is determined by the degree of its membership with a value range from 0 to 1. The advantages of this algorithm include being able to cluster more than one variable at once [15]. The purpose of the cluster using Fuzzy C-Means is to get the center of the cluster which will later be used to find out the data that enters a cluster.

3. *Genetic Algorithm Optimization*

A genetic algorithm is an algorithm based on Charles Darwin's theory of evolution, this algorithm is often used as a tool to solve a problem. This algorithm reflects the natural selection process in which the most suitable individuals are selected for reproduction to produce the next generation. Genetic algorithms provide optimal solutions with various objective variations and have many advantages compared to other algorithms when viewed in terms of their capabilities [16].

The advantage of this genetic algorithm is that it can solve problems in engineering and science that cannot be solved by ordinary deterministic algorithms even with polynomially increased time. In general, this genetic algorithm has pseudocode which can be seen in Figure 3.

Genetic Algorithm

```
Begin
t←0
initialize P(t)
evaluate P(t)
while (not termination condition)
do
 begin
   recombine P(t) to yield C(t)
   evaluate C(t)
   select P(t+1) from P(t) and C(t)
   t ← t+1
 end
end
```

Fig. 3. Pseudocode Genetic Algorithm

Genetic Algorithm will be inserted in K-Means and FCM. The insertion process carried out on the K-Means algorithm is in the initial initialization section of the cluster center, while in the Fuzzy C-Means it is for the initial determination of the cluster membership value.

Hybrid Genetic Algorithm – K-Means

Genetic Algorithm in K-Means clustering is often used as an optimization because k means clustering has a weakness in the initialization stage of the center value or centroid [17]. Where the initial selection of the initial centroid on the K-means is done randomly so that it can affect the results of the clustering. The genetic algorithm process in determining the cluster center point with fast performance is often called the Fast Genetic K-means Algorithm which is in the flowchart in Figure 4.

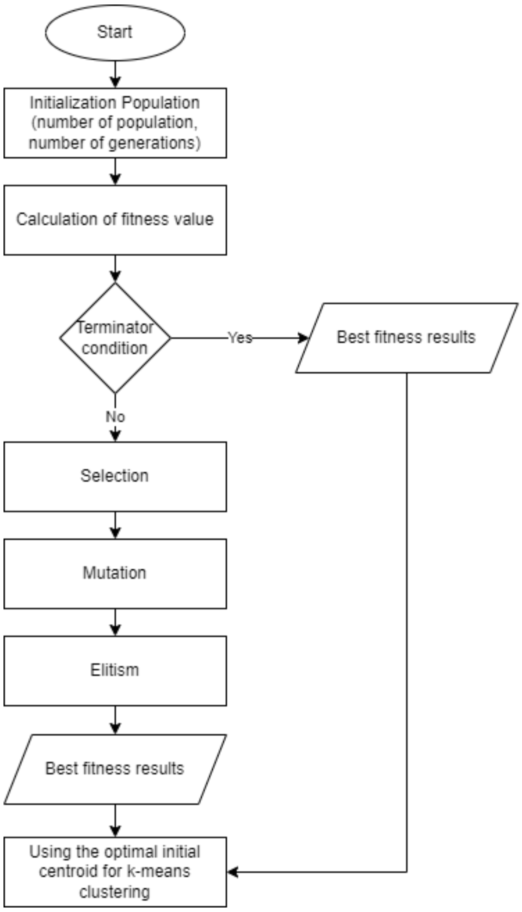


Fig. 4. Genetic Algorithm -K Means Flowchart

The steps for the hybrid Genetic Algorithm and K-Means are as follows [18]:
1) Determining the parameters of the Genetic Algorithm
2) Determine a random number as many as the population as a candidate for the center of the cluster
3) Calculate the minimum distance and fitness of each cluster center
4) Carry out the selection, crossover, and mutation process
5) Repeat steps 3 and 4 until stable

Hybrid Algoritma Genetika – FCM

In the Fuzzy C-Means algorithm, the Genetic Algorithm is used to generate cluster center candidates by initializing the value of the degree of cluster membership so that it is not done randomly. The research steps of the genetic algorithm fuzzy clustering c-means can be seen in Figure 5 [19].

The steps for the hybrid Genetic Algorithm and Fuzzy C-Means are as follows [20]:
1) Determining the parameters of the Genetic Algorithm
2) Determine a random number as many as the population as a candidate for the center of the cluster
3) Calculate the degree of membership and fitness of each cluster center.
4) Carry out the selection, crossover, and mutation process.

Repeat steps 3 and 4 until stable

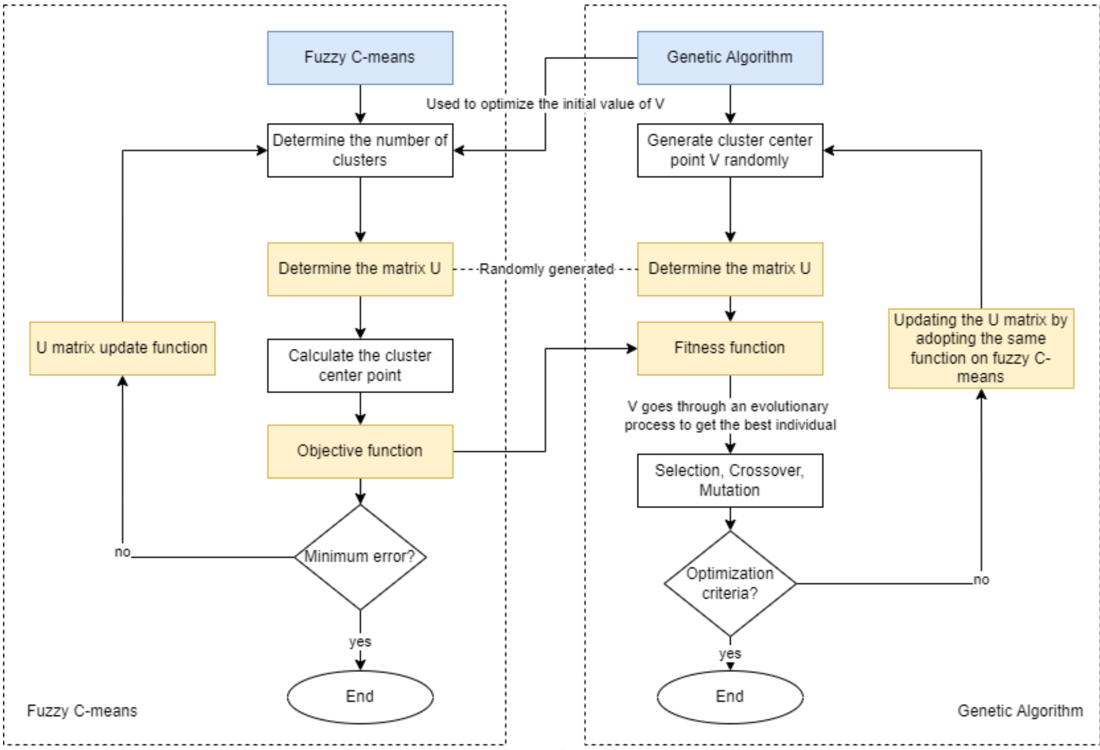


Fig. 5. Genetic Algorithm Flowchart - FCM

Results and Discussion

In this study, there are 2 experimental models carried out, namely clustering using the original algorithm without optimization with the optimized algorithm. Model 1 is K-Means only with Hybrid Genetic Algorithm - K-Means and model 2 is Fuzzy C-Means only with Hybrid Genetic Algorithm - Fuzzy C-Means. Each model will be recorded the necessary iterations until it converges. In model 1 each experiment was carried out as many as 6 iterations, while in model 2 it was carried out as many as 10 iterations.

The experiment conducted using model 1 is an experiment using the K-Means clustering method. In this case, 3 K clusters are used according to the level of the existing question categories, namely difficult (C0), medium (C1), and easy (C2). The results of the experiment using the original K-Means method without any optimization using the initial value of the cluster center (centroid) at random can be seen in Table 2.

Table 2. K-Means Model Cluster Results

Iteration	Cluster center (Centroid)			Cluster Result		
	C0	C1	C2	C0	C1	C2
1	[6,27]	[31,2]	[27,6]	26	4	20
2	[5,28]	[32,1]	[24,9]	26	4	20
3	[5,28]	[32,1]	[24,9]	26	4	20
4	[5,28]	[26,7]	[23,10]	26	6	18
5	[4,29]	[29,4]	[21,12]	25	12	13
6	[4,29]	[29,4]	[21,12]	25	12	13

Based on the results in Table 2, it can be seen that the value of the cluster center (centroid) in iterations 1 to 4 is still moving each other and the centroid is no longer moving (converging) when entering iterations 5 to 6 with the centroid value at C0 4.29; C1 29.4; and C2 21.12. From the results, the K-Means of 50 questions resulted in a cluster of questions with a Difficult level (C0) of 25 questions, Medium (C1) as many as 12 questions, and Easy (C2) as many as 13 questions. Members of the questions that are in each cluster generated from the K-Means model are as follows:

1) C0 number of questions: 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 18, 20, 22, 23, 25, 26, 27, 28, 29, 32, 34, 35, 36, 37 and 39
2) C1 number of questions: 4, 8, 17, 38, 40, 41, 43, 44, 46, 48, 49 and 50
3) C2 number of questions: 1, 2, 3, 9, 19, 21, 24, 30, 31, 33, 42, 45 and 47

Subsequent experiments on Model 1 were also carried out using K-Means which were optimized using the Genetic Algorithm. Genetic Algorithm in K-Means clustering is used to optimize the determination of the initial centroid to have better results than using a random centroid determination. The results of the experiment using the Hybrid Genetic Algorithm - K-Means method can be seen in Table 3.

Table 3. Cluster Results of Genetic Algorithm Model - K-Means

Iteration	Pusat Cluster (Centroid)			Hasil Cluster		
	C0	C1	C2	C0	C1	C2
1	[4,29]	[19,14]	[28,5]	23	14	13
2	[4,29]	[19,14]	[27,6]	25	8	17
3	[4,29]	[20,13]	[28,5]	25	12	13
4	[4,29]	[20,13]	[28,5]	25	12	13
5	[4,29]	[20,13]	[28,5]	25	12	13
6	[4,29]	[20,13]	[28,5]	25	12	13

In Table 3 it can be seen that the results of the genetic algorithm - K-Means cluster model in iterations 1 to 2, the value of the center of the cluster (centroid) is still moving and the centroid is no longer moving (converging) when entering iterations 3 to 6 with the centroid value at C0. 4.29; C1 20.13; and C2 28.5. From the results, the K-Means of 50 questions resulted in a cluster of questions with a Difficult level (C0) of 25 questions, Medium (C1) as many as 12 questions, and Easy (C2) as many as 13 questions. The members of the questions that are in each cluster generated from the Genetic Algorithm - K-Means model are as follows:

1) C0 question number: 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 18, 20, 22, 23, 25, 26, 27, 28, 29, 32, 34, 35, 36, 37 and 39
2) C1 question number 4, 8, 17, 38, 40, 41, 43, 44, 46, 48, 49 and 50
3) C2 question number 1, 2, 3, 9, 19, 21, 24, 30, 31, 33, 42, 45 and 47

From the results of the clustering experiment using the 1 K-Means model only with the Genetic Algorithm - K-Means, it can be seen that by using the same number of K clusters and iterations, the Genetic Algorithm - K-Means model reaches convergence faster when entering the 3rd iteration compared to with the usual K-Menas which only reached convergence when it entered iteration 5. In addition, it can be seen that the determination of the initial centroid has an influence on the process and results of the existing clusters. In addition, the cluster center owned by the K-Means model only with the Genetic Algorithm - K-Means also has differences which can be seen in Table 4.

Table 4. Comparison of Cluster Results Model 1

Method	Cluster Center (Centroid)			Cluster result		
	C0	C1	C2	C0	C1	C2
K-Means	[4,29]	[29,4]	[21,12]	23	14	13
Genetic Algorithm – K-Means	[4,29]	[20,13]	[28,5]	25	8	17

Table 4 shows the results of the comparison of model 1 clusters when iterations have converged. From the existing results, the center of the cluster has a difference in its C1 and C2 centroids. Even so, the results of the clusters that exist between the K-Means method and the Genetic Algorithm - K-Means have the same results, namely from the 50 questions that result in a cluster of questions with a Difficult level (C0) of 25 questions, Medium (C1) as many as 12 questions. and Easy (C2) as many as 13 questions. Members of the questions that are in each cluster resulting from model 1 are as follows:

1) C0 number of questions: 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 18, 20, 22, 23, 25, 26, 27, 28, 29, 32, 34, 35, 36, 37 and 39
2) C1 number of questions: 4, 8, 17, 38, 40, 41, 43, 44, 46, 48, 49 and 50
3) C2 number of questions 1, 2, 3, 9, 19, 21, 24, 30, 31, 33, 42, 45 and 47

The experiment conducted using model 2 is an experiment using the Fuzzy C-Means clustering method. In this case, the K cluster used is also the same as model 1 as many as 3 according to the level of the existing question categories, namely difficult (C0), medium (C1), and easy (C2). There are additional parameters used in the clustering process using Fuzzy C-Means, namely the value of the weighting power of 2, the maximum iteration of 10, and the expected epsilon or error approaching 0.001. The results of the experiment using the original Fuzzy C-Means method without any optimization using the initial value of the membership degree value at random can be seen in Table 5.

Table 5. Fuzzy C-Means Cluster Results

Iteration	Cluster center (Centroid)			Objective Function Difference	Cluster result		
	C0	C1	C2		C0	C1	C2
1	[14.7,18.3]	[14.3,18.7]	[14.2,18.8]	5.2732	24	13	13
2	[15.6,17.4]	[14,19]	[13.7,19.3]	1.1975	13	11	26
3	[18.5,14.5]	[13,20]	[11.5,21.5]	0.3395	24	1	25
4	[22.6,10.4]	[9.7,23.3]	[6.8,26.2]	1.3385	25	1	24
5	[24.9,8.1]	[9,24]	[4.4,28.6]	1.2665	24	3	23
6	[25.6,7.4]	[9.6.23.4]	[3.6,29.4]	0.1835	23	6	21
7	[25.8,7.2]	[10.1,22.9]	[3.3,29.7]	0.3052	20	8	22
8	[26,7]	[10.6,22.4]	[3.2,29.8]	1.7582	23	6	21
9	[25.9,7.1]	[8.7,24.3]	[5,28]	2.4323	22	7	21
10	[24.7,8.3]	[4,29]	[6.3,26.7]	2.4310	24	26	0

Based on the results in Table 5, it can be seen that the value of the cluster center (centroid) in iterations 1 to 10 is still moving and not convergent. Judging from the results of the difference in the objective function, iteration 6 has the smallest number of 0.1835 with a centroid value at C0 of 25.6.7.4; C1 9.6,23.4; and C2 3.6,29.4.

From the results, the Fuzzy C-Means of 50 questions resulted in a cluster of questions with a Difficult level (C0) of 23 questions, Medium (C1) as many as 6 questions and Easy (C2) as many as 21 questions. Members of the questions that are in each cluster generated from the Fuzzy C-Means model are as follows:

1) C0 question number 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 18, 20, 22, 23, 25, 26, 27, 28, 29, 32, 34, 35, 36, 37 and 39
2) C1 question number 48, 15, 17, 18, 22, 23, 38 and 40
3) C2 question number 1, 2, 3, 4, 9, 19, 21, 24, 30, 31, 33, 41, 42, 43, 44, 45, 46, 47, 48, 49 and 50

Subsequent experiments on Model 2 were also carried out using Fuzzy C-Means which were optimized using Genetic Algorithms. Genetic Algorithm in Fuzzy C-Means clustering is used to optimize the initial determination of the value of the degree of cluster membership to have better results than using the determination of the degree of cluster membership at random. The results of the experiment using the Hybrid Genetic Algorithm - Fuzzy C-Means method can be seen in Table 6.

Table 6. Genetic Algorithm Cluster Results - Fuzzy C-Means

Iteration	**Cluster center (Centroid)**			**Objective Function Difference**	**Cluster result**		
	C0	**C1**	**C2**		**C0**	**C1**	**C2**
1	[8.2,24.8]	[16.8,16.2]	[21.3,11.7]	4.1105	25	12	13
2	[5.2,27.8]	[20.9,12.1]	[25, 8]	2.6505	24	12	14
3	[4.3,28.7]	[20.7,12.3]	[27.2,5.8]	0.7475	25	8	17
4	[4.1,28.9]	[20.6,12.4]	[28,5]	0.7085	25	12	13
5	[4.3,28.7]	[11.4,21.6]	[26.2,6.8]	1.0374	27	0	23
6	[3.7.,29.3]	[12.5,20.5]	[26.2,6.8]	1.3107	22	7	21
7	[3.6,29.4]	[13.6,19.4]	[26.3,6.7]	1.3110	23	6	21
8	[3.6,29.4]	[14.9,18.1]	[26.5,6.5]	0.0114	23	6	21
9	[3.4,29.6]	[10,23]	[26.7,6.3]	2.8593	21	14	15
10	[9.7,23.3]	[19,14]	[27.2,5.8]	2.3944	39	11	0

In Table 6, it can be seen that the results of the Genetic Algorithm - Fuzzy C-Means cluster model in iterations 1 to 10, the value of the cluster center (centroid) is also still moving and the centroids are not convergent. From the results of the difference in the objective function, iteration 8 has the smallest number of 0.0114 with a centroid value at C0 of 3.6,29.4; C1 14.9,18.1; and C2 26.5,6.5. From the results, the Fuzzy C-Means of 50 questions resulted in a cluster of questions with a Difficult level (C0) of 23 questions, Medium (C1) as many as 6 questions, and Easy (C2) as many as 21 questions. Members of the questions that are in each cluster generated from the Genetic Algorithm - Fuzzy C-Means model are as follows:

1) C0 question number: 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 18, 20, 22, 23, 25, 26, 27, 28, 29, 32, 34, 35, 36, 37 and 39
2) C1 question number 8, 15, 17, 18, 22, 23, 38 and 40
3) C2 question number 1, 2, 3, 4, 9, 19, 21, 24, 30, 31, 33, 41, 42, 43, 44, 45, 46, 47, 48, 49 and 50

From the results of the clustering experiment using only 2 Fuzzy C-Means models with the Genetic Algorithm - Fuzzy C-Means, it can be seen that by using the same number of K clusters and iterations, the 2 models have not reached convergence because of the difference in the existing objective function is still greater than the expected error value (epsilon) is 0.0001. In addition, it can be seen that the initial determination of the value of the degree of cluster membership has an influence on the process and results of the existing cluster. In addition, from the results of the difference in the smallest objective function which is owned by the Fuzzy C-Means model only with the Genetic Algorithm - Fuzzy C-Means also has a difference which can be seen in Table 7.

Table 7. Comparison of Cluster Results Model 2

Method	**Cluster center (Centroid)**			**Cluster result**		
	C0	**C1**	**C2**	**C0**	**C1**	**C2**
Fuzzy C-Means	[25.6,7.4]	[9.6.23.4]	[3.6,29.4]	23	6	21
Genetic Algorithm – *Fuzzy C-Means*	[3.6,29.4]	[14.9,18.1]	[26.5,6.5]	23	6	21

Table 7 shows the results of the comparison of model 2 clusters from the results of the smallest difference in objective functions. From the results, the cluster center has differences in all the centers of the C0, C1, and C2 centroids. Even so, the results of the clusters that exist between the Fuzzy C-Means method and the Genetic Algorithm - Fuzzy C-Means have the same results, namely from 50 questions that produce a cluster of questions with a Difficult level (C0) of 23 questions, Medium (C1) of 6 questions and Easy (C2) as many as 21 questions. Members of the questions that are in each cluster resulting from model 2 are as follows:

1) C0 question number: 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 18, 20, 22, 23, 25, 26, 27, 28, 29, 32, 34, 35, 36, 37 and 39
2) C1 question number 8, 15, 17, 18, 22, 23, 38 and 40

C2 question number 1, 2, 3, 4, 9, 19, 21, 24, 30, 31, 33, 41, 42, 43, 44, 45, 46, 47, 48, 49 and 50

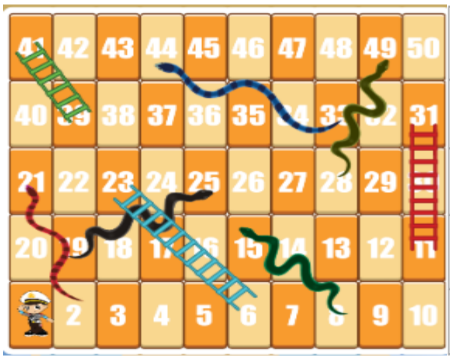
Table 8. Comparison of Cluster Results of All Models

Model	**Cluster question result**		
	C0	**C1**	**C2**
1	5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 18, 20, 22, 23, 25, 26, 27, 28, 29, 32, 34, 35, 36, 37 and 39	4, 8, 17, 38, 40, 41, 43, 44, 46, 48, 49 and 50	1, 2, 3, 9, 19, 21, 24, 30, 31, 33, 42, 45 and 47
2	5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 18, 20, 22, 23, 25, 26, 27, 28, 29, 32, 34, 35, 36, 37 and 39	8, 15, 17, 18, 22, 23, 38 dan 40	1, 2, 3, 4, 9, 19, 21, 24, 30, 31, 33, 41, 42, 43, 44, 45, 46, 47, 48, 49 and 50

Based on the results in Table 8, it can be seen that for the results of the question cluster for Difficult (C0) the results of the two models are the same, namely 25 with the same number of questions. Meanwhile, the Medium (C1) and Easy (C2) question clusters have differences both in terms of number and number of questions. From the results obtained, the next process is the implementation of the results of the difficulty level cluster on the LantaSmart game board.

As shown in Figure 6 (a), the LantaSmart board game which resembles a snake and ladder has 50 boxes, of which 3 boxes have ladders placed in them (boxes 6, 11, and 39) for difficult questions (C0), 5 boxes which have the placement of snakeheads (15, 21, 25, 44 and 49) for easy questions (C2), and the remaining 42 boxes are boxes that do not have ladder or snake placement for medium questions (C1). From the results of the existing clusters, it can be seen that Model 1 Genetic Algorithm - K-Means has better results. Model 2 Genetic Algorithm Fuzzy C-Means. In implementing this question, using the help of ranking the existing test results due to differences in the board designed with the results of the existing question cluster.

There are 25 difficult questions (C0), while on the board there are only 3 question boxes. There are 12 medium questions (C1), while the board has 42 boxes that can be filled in. There are 13 difficult questions (C2), while the board only has 5 question boxes. In the ranking results, the board containing the stairs up for 3 difficult questions will be filled with questions with question numbers 13, 22, and 25. For the board containing the swap heads for 5 easy questions, the number 2 will be filled in. 3, 19, 21, and 30. On boards that do not have ladders or snakes placed for moderate questions (C1), the remaining number of questions will be filled in the order by the results of the ranking of questions from the largest to the smallest value with the assumption that it is getting closer to the finish in box 50 the level of the questions will be more difficult as shown in Figure 6.



23	16	34	21	26	29	6	12	30	27
39	25	35	20	11	10	7	37	32	28
3	40	8	17	19	18	15	5	36	14
38	49	50	44	41	2	48	46	43	4
24	31	33	42	45	13	47	1	9	22

(a) (b)

Fig. 6. (a) LantaSmart Board Game, (b) Implementing Question Number Arrangements on the LantaSmart Board Game

So, from the research that has been done, it can be seen that clustering the level of difficulty of questions for the development of the LantaSmart educational game can be done using clustering techniques, namely K-means and Fuzzy C-Means which can also be optimized using genetic algorithms to get better results.

Conclusions

Genetic Algorithm can be applied to optimize the K-means and Fuzzy C-Means clustering algorithms. Optimization is done to speed up convergence or in other words, reduce the number of iterations. The genetic algorithm acts as a substitute for the role of finding the center of the cluster. The K-Means algorithm can be used for the initial initialization of the cluster center, while the Fuzzy C-Means algorithm is used to determine the initial cluster membership value so that it is not determined randomly. Genetic Algorithm Optimization K-Means speeds up the convergence process in the 3rd iteration when compared to the usual K-Means in the 5th iteration.

References

[1] J. Shi, R. Renwick, N. E. Turner, and B. Kirsh, "Understanding the lives of problem gamers: The meaning, purpose, and influences of video gaming," *Comput. Human Behav.*, vol. 97, pp. 291–303, Aug. 2019, doi: 10.1016/j.chb.2019.03.023.

[2] V. S. Zirawaga, A. I. Olusanya, and T. Maduku, "Gaming in Education: Using Games as a Support Tool to Teach History," *J. Educ. Pract.*, vol. 8, no. 15, pp. 55–64, 2017.

[3] N. Pellas, P. Fotaris, I. Kazanidis, and D. Wells, "Augmenting the learning experience in primary and secondary school education: a systematic review of recent trends in augmented reality game-based learning," *Virtual Real.*, vol. 23, no. 4, pp. 329–346, Dec. 2019, doi: 10.1007/s10055-018-0347-2.

[4] S. Amami Pramuditya, M. S. Noto, and D. Syaefullah, "Game Edukasi Rpg Matematika," *Eduma Math. Educ. Learn. Teach.*, vol. 6, no. 1, p. 77, 2017, doi: 10.24235/eduma.v6i1.1701.

[5] C. Dichev and D. Dicheva, "Gamifying education: what is known, what is believed and what remains uncertain: a critical review," *Int. J. Educ. Technol. High. Educ.*, vol. 14, no. 1, p. 9, Dec. 2017, doi: 10.1186/s41239-017-0042-5.

[6] J. Qi, Y. Yu, L. Wang, J. Liu, and Y. Wang, "An effective and efficient hierarchical K -means clustering algorithm," *Int. J. Distrib. Sens. Networks*, vol. 13, no. 8, p. 155014771772862, Aug. 2017, doi: 10.1177/1550147717728627.

[7] M. Anusha and J. G. . Sathiaseelan, "An Enchanced K-Means Genetic Algoritms for Optimal Clustering," *IEEE Int. Conf. Comput. Comput. Res.*, vol. 18, no. 11, pp. 1463–1476, 2014, doi: 10.1068/a181463.

[8] O. Mohammadrezapour, O. Kisi, and F. Pourahmad, "Fuzzy c-means and K-means clustering with genetic algorithm for identification of homogeneous regions of groundwater quality," *Neural Comput. Appl.*, vol. 32, no. 8, pp. 3763–3775, Apr. 2020, doi: 10.1007/s00521-018-3768-7.

[9] A. Rachman, D. A. Fatimah, H. Nugroho, Sulistyowati, W. Widodo, and W. A. Surasmi, "Development of educational traffic signs game based on android for elementary school students," in *Emerging Perspectives and Trends in Innovative Technology for Quality Education 4.0*, Routledge, 2020, pp. 166–169.

[10] S. Ismail, S. Abdennadher, and W. Abouelsaadat, "Rules on wheels: A serious game for teaching traffic signs," *2016 8th Int. Conf. Games Virtual Worlds Serious Appl. VS-Games 2016*, 2016, doi: 10.1109/VS-GAMES.2016.7590360.

[11] M. S. Riaz, A. Cuenen, D. Janssens, K. Brijs, and G. Wets, "Evaluation of a gamified e-learning platform to improve traffic safety among elementary school pupils in Belgium," *Pers. Ubiquitous Comput.*, vol. 23, no. 5–6, pp. 931–941, Nov. 2019, doi: 10.1007/s00779-019-01221-4.

[12] R. O. R. I. Ophelia, "The Influence of Traffic Violation to Accident Potential on Young Motorcyclists," 2016.

[13] S. Soehodho, "Public transportation development and traffic accident prevention in Indonesia," *IATSS Res.*, vol. 40, no. 2, pp. 76–80, Jan. 2017, doi: 10.1016/j.iatssr.2016.05.001.

[14] S. Wikaisuksakul, "A multi-objective genetic algorithm with fuzzy c-means for automatic data clustering," *Appl. Soft Comput.*, vol. 24, pp. 679–691, Nov. 2014, doi: 10.1016/j.asoc.2014.08.036.

[15] H. Li, "Application of K-means clustering algorithm in the analysis of college students' online entertainment consumption," *J. Phys. Conf. Ser.*, vol. 1570, no. 1, p. 012018, Jun. 2020, doi: 10.1088/1742-6596/1570/1/012018.

[16] W. Rankothge, F. Le, A. Russo, and J. Lobo, "Optimizing Resource Allocation for Virtualized Network Functions in a Cloud Center Using Genetic Algorithms," *IEEE Trans. Netw. Serv. Manag.*, vol. 14, no. 2, pp. 343–356, Jun. 2017, doi: 10.1109/TNSM.2017.2686979.

[17] A. Sinha and P. K. Jana, "A hybrid MapReduce-based k-means clustering using genetic algorithm for distributed datasets," *J. Supercomput.*, vol. 74, no. 4, pp. 1562–1579, Apr. 2018, doi: 10.1007/s11227-017-2182-8.

[18] M. A. El-Shorbagy, A. Y. Ayoub, I. M. El-Desoky, and A. A. Mousa, "A Novel Genetic Algorithm Based k-means Algorithm for Cluster Analysis," 2018, pp. 92–101.

[19] L. Jie, W. Liu, Z. Sun, and S. Teng, "Hybrid fuzzy clustering methods based on improved self-adaptive cellular genetic algorithm and optimal-selection-based fuzzy c-means," *Neurocomputing*, vol. 249, pp. 140–156, Aug. 2017, doi: 10.1016/j.neucom.2017.03.068.

[20] G. M. Lee and X. Gao, "A Hybrid Approach Combining Fuzzy c-Means-Based Genetic Algorithm and Machine Learning for Predicting Job Cycle Times for Semiconductor Manufacturing," *Appl. Sci.*, vol. 11, no. 16, p. 7428, Aug. 2021, doi: 10.3390/app11167428.