Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2023. Vol. 3. No. 1; e-ISSN : 2808-4519

Grouping Number of Library Members For Determining the Location of Socialization Using Clustering Method

Sella Dwi Pratiwi [1] , Achmad Fauzi [2] , I Gusti Prahmana [3]

[1,2,3] Information Systems, STMIK KAPUTAMA
Jl. Veterans No. 4A-9A, Binjai, North Sumatra, Indonesia
selladwipratiwi623@gmail.com [1] , fauzie.kaputama@gmail.com [2] , igustiprahmana4@gmail.com [3]*

Abstract

The high use of smartphones at this time led to a decline in public interest in reading books in the library directly. Especially students and students. This is certainly a problem for the Langkat Regency Archives and Libraries Office. Socialization is needed to increase efforts to read interest in the community. The right socialization location must have several criteria so that the socialization carried out is right on target. The existence of a database for each member of the library will facilitate the location selection process. Data mining techniques can classify the number of library members based on the results of large data analysis into information in the form of patterns. The clustering method is a method in data mining that can analyze data with the aim of grouping data based on the same characteristics. The K-Means algorithm is a simple algorithm for classifying a large number of objects with certain attributes into clusters which are usually used in data mining.

Keywords: *Clustering, Data Mining, K-Means, Outreach.*

1. Introduction

The library is a place to access information in any format, both a collection of printed books, and digital (in the form of data that can be accessed via a computer network). The role of the library seeks to maintain and improve the efficiency and effectiveness of the teaching and learning process which aims to help people of all ages to acquire various fields of knowledge.

The Langkat Regency Archives and Library Service is a Langkat Regency government agency that provides various reading books and other sources of knowledge information. Every member of the public, students, university students and researchers who wish to seek information and knowledge or references, can obtain it from the Archives and Libraries Office of Langkat Regency. But the problems that occur in the era of increasingly sophisticated technology make the library office a lack of interest in reading by the public to get information, this is because many people are lazy to visit the library, which is caused by the increasing use of smartphones . Where with today's *smartphones it is very easy for people to* find information, people can get it with just one grip without even having to leave the house . For this reason, the Langkat Regency Archives and Library Service must make efforts to carry out socialization to increase efforts to read interest in the community. In determining the location of course must have several criteria in determining the appropriate location of socialization.

The Archives and Libraries Service has *a database* of every registered member and member who frequently visits the library. From this data, it can be processed and used as material for consideration in determining the right socialization location for the community to increase reading interest. To obtain precise location information, it is necessary to use data mining techniques in classifying the number of library members who frequently visit and it can be known which sub-districts rarely visit the library. Thus the Archives and Libraries Office can obtain the right socialization location. Data mining techniques are used because they analyze large data into information in the form of patterns that have meaning for decision support. Supported by applying the *Clustering method* , where this method produces a high level of similarity within one class and a low level of similarity between classes. The similarity in question is a measurement in terms of numeric for two objects. The similarity value between the two objects will be higher if the two objects being compared have a high similarity [1].

From a journal entitled "Data Mining Grouping Electricity Payment Using Clustering Method: Case Study of the Binjai Post Office". From this study it was concluded that the grouping of electricity payments which were divided into 3 clusters resulted in information on cluster 1, the electricity payment group in the Power group (X) was 900 VA, for the Total Payment group (Y) was 200,001-400,000 and in the Fines group (Z) that had to be issued was 3,000 per month. In cluster 2 the electricity payment group in the Power group (X) is 450 VA,

for the Total Payment group (Y) is 200,001-400,000 and in the group The fine (Z) that must be issued is 5,000 per month. In cluster 3 groups electricity payment in the Power group (X) is 1,300 VA, for the Total group Payment (Y) is 600,001-800,000 and in the Fines group (Z) that must be paid issue is 50,000 per month [2] . In research entitled "Application of Data Mining to Determine the Location of New Student Admissions Promotion at the University of Banten Jaya (K-Means Clustering Method)". From this study it was concluded that using K-Means *Clustering* was formed into three clusters, the first cluster was in the high category, the second cluster was in the low category and the three clusters were in the medium category. A promotion strategy for prospective new students that is right on target for each region based on clusters is formed [3] . Based on the journal entitled "Group the Number of Library Visitors in DKI Jakarta Using the K-Means Method", it concluded that the libraries that had the most visits were KPAK Pulau 1000 (RPTRA), KPAK Jakarta Pusat (Pusling) and Cikini, which were dominated by visitors and users. The average number of visitors who come is 87,180 [4].

2. Problem Solving Methodology

2.1. Data Mining

The process of discovering knowledge in databases is often referred to as data mining or data mining. Data mining is an analysis step of a set of data that is generally large in size to obtain relationships between these data and summarize them in a form that is easy to understand and use. The resulting relationships and summaries are generally in the form of *models* or *patterns* .

Data mining is the extraction of important or interesting information or patterns from data in large *databases.* In scientific journals, data mining is also known as Knowledge Discovery in *Database* (KDD). Broadly speaking, KDD includes three stages, namely *preprocessing, process* (data mining) and *post processing* . In conclusion, data mining is the process of finding certain patterns and a large data or database to obtain useful information. Types or data mining techniques applied are *association, classification* , and *clustering* . *Associations* look for patterns of relationships found in data or databases, *classification* , and *clustering* are more likely to find patterns for grouping [5] .

2.2. Clustering

Clustering is a data analysis method, which is often included as a data mining method, the purpose of which is to group data with the same characteristics.

Clustering is a grouping method based on a measure of proximity or (similarity). Clustering is different from groups, if the group means the same group, condition if not definitely not the group. However, the clusters do not have to be the same only groups based on closeness and a characteristic of the existing sample, namely by using the *Eucledian distance formula* [6].

The main objective of the *clustering method* is to group a number of data/objects into *clusters* (groups) so that each *cluster* contains data that is as similar as possible. The *clustering* method seeks to place similar (closely spaced) objects in one group and make the distance between groups as far as possible. This means that objects in one group are very similar to each other and different from objects in other groups.

2.3. *K-Means* Algorithm

K-Means algorithm is a relatively simple algorithm for classifying or grouping a large number of objects with certain attributes into groups (*clusters*) of K. In the *K-Means algorithm* , the number of K *clusters* is predetermined.

K-Means is the most well-known clustering method and is widely used in various fields because it is simple, easy to implement, has the ability to cluster large data, is able to handle outlier data, and the time complexity is linear with the number n being the number of documents, K being the number of clusters, and T being the number of iterations. K-means is a partitioning clustering method that separates data into different groups. By partitioning iteratively, K-Means is able to minimize the average distance of each data to its cluster [7].

2.4. Socialization

Socialization is a process of teaching and learning in behaving in society. Some people also say that socialization is the process of instilling values, habits, and rules of behavior in society from one generation to another. In the process of socialization itself, humans are adapted to their respective social roles and statuses within social groups. With the socialization process, a person can know, understand and at the same time carry out their rights and obligations based on their respective status roles according to the culture of society. Furthermore, in the process of recognizing the rights and obligations of an adult human being, every individual or human needs to socialize to learn and develop patterns of social behavior with other members of society.

3. Method Application

In using the *clustering method*, the initial process carried out to form clusters is to transform data into *numeric form* with predetermined codes, then determine the number of groups (K), calculate the centroids, calculate the distance of objects to the *centroids* and then group them based on the closest distance, if no objects move or group then the iteration is complete. To determine the group of an object, the first thing to do is measure the Euclidean distance between two object points (X, Y and Z).

Table 1: Education

Code	Education (X)
1	Kindergarten/PAUD
2	SD
3	JUNIOR HIGH SCHOOL
4	High School / Equivalent
5	D1
6	D3
7	S1
8	S2
9	S3

Table 2: Occupation

Code	Occupation (Y)
1	Student
2	Student
3	Teacher
4	Lecturer
5	Government employees
6	Private employees
7	Retired
8	Self-employed
9	Other

Table 3: District

Code	District (Z)
1	Babalan
2	Bahorok
3	Attack Bar
4	Besitang
5	Binjai
6	West Brandan
7	Gebang
8	humiliate
9	Kuala
10	Kutambaru
11	Tualang Field
12	Dairy Base
13	Override
14	A cup
15	Sei Bingai
16	Sei Lepan
17	Finished
18	Seraph
19	stab
20	Tanjung Pura
21	Wampu

Then transform the criteria data above to be calculated using the *clustering method* . And the formula for performing the calculation is as follows:

$$d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \ldots + (x_{ki} - k_j)^2}$$

Data transformation of the data above can be seen in the table below.

Table 4: Transformation Data

No	object	X	Y	Z
1	A	4	6	14
2	B	4	6	14
3	C	7	2	5
4	D	3	6	14
5	E	4	6	14
6	F	4	9	11
7	G	4	3	19
8	H	7	3	21
9	I	4	9	19
10	J	2	6	5
11	K	7	8	21
12	L	7	9	5
13	M	7	3	14
14	N	3	6	21
15	O	3	6	19
16	P	3	6	19
17	Q	7	2	19
18	R	7	3	20

No	object	X	Y	Z
19	S	7	3	19
20	Q	3	6	19

Then form *a cluster* into 3 groups (K=3) and determine the *centroid center point* . The *clustering* calculation process is as follows.
K=3 *Centroids*
C_1 = (7, 3, 21) taken randomly from H data
C_2 = (4, 9, 19) taken randomly from data I
C_3 = (2, 6, 5) taken randomly from data J

Information :
The initial centroid was determined randomly.
Then do the calculations like the calculation process below:
Iteration 1 :
1.A (4,6,14)
C1 = (7, 3, 21) = √((4-7)^2+(〖 6-3) 〗 ^2+(14-21)²=) 8,19
C2 = (4, 9, 19) = √((4-4)^2+(〖 6-9) 〗 ^2+(14-19)²=) 5.83
C3 = (2, 6, 5) = √((4-2)^2+(〖 6-6) 〗 ^2+(14-5)^2 =) 9,22
2. B (4,6,14)
C1 = (7, 3, 21) = √((4-7)^2+(〖 6-3) 〗 ^2+(14-21)²=) 8,19
C2 = (4, 9, 19) = √((4-4)^2+(〖 6-9) 〗 ^2+(14-19)²=) 5.83
C3 = (2, 6, 5) = √((4-2)^2+(〖 6-6) 〗 ^2+(14-5)^2 =) 9,22
3. C(7,2,5)
C1 = (7, 3, 21) = √((7-7)^2+(〖 2-3) 〗 ^2+(5-21)²=) 16.03
C2 = (4, 9, 19) = √((7-4)^2+(〖 2-9) 〗 ^2+(5-19)²=) 15.94
C3 = (2, 6, 5) = √((7-2)^2+(〖 2-6) 〗 ^2+(5-5)^2 =) 6.40
4. D (3,6,14)
C1 = (7, 3, 21) = √((3-7)^2+(〖 6-3) 〗 ^2+(14-21)²=) 8.60
C2 = (4, 9, 19) = √((3-4)^2+(〖 6-9) 〗 ^2+(14-19)²=) 5.92
C3 = (2, 6, 5) = √((3-2)^2+(〖 6-6) 〗 ^2+(14-5)^2 =) 9.96
5.E (4,6,14)
C1 = (7, 3, 21) = √((4-7)^2+(〖 6-3) 〗 ^2+(14-21)²=) 8,19
C2 = (4, 9, 19) = √((4-4)^2+(〖 6-9) 〗 ^2+(14-19)²=) 5.83
C3 = (2, 6, 5) = √((4-2)^2+(〖 6-6) 〗 ^2+(14-5)^2 =) 9,22

From the calculation above, the results of the iteration 1 calculation are obtained, which are as shown in the table below.

Table 5: Results of Iteration 1

No	object	X	Y	Z	C_1	C_2	C_3	Group
1	A	4	6	14	8,19	5.83	9,22	2
2	B	4	6	14	8,19	5.83	9,22	2
3	C	7	2	5	16.03	15.94	6,40	3
4	D	3	6	14	8.60	5.92	9.06	2
5	E	4	6	14	8,19	5.83	9,22	2
6	F	4	9	11	12.04	8	7	3
7	G	4	3	19	3.61	6	14,46	1
8	H	7	3	21	0	7	17.03	1
9	I	4	9	19	7	0	14,46	2
10	J	2	6	5	17.03	14,46	0	3
11	K	7	8	21	5	3.74	16.88	2
12	L	7	9	5	17.09	14,32	5.83	3
13	M	7	3	14	7	8.37	10.72	1
14	N	3	6	21	5	3.74	16.03	2
15	O	3	6	19	5,39	3,16	14.04	2
16	P	3	6	19	5,39	3,16	14.04	2
17	Q	7	2	19	2,24	7,62	15.39	1
18	R	7	3	20	1	6,78	16.09	1
19	S	7	3	19	2	6,71	15,17	1
20	Q	3	6	19	5,39	3,16	14.04	2

After calculating using the existing *cluster formula, the groups based on the minimum distance to the nearest centroid* are:

Old Groups:	(0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0)
New Groups:	(2	2	3	2	2	3	1	1	2	3	2	3	1	2	2	2	1	1	1	2)

There is a group change, followed by the following iteration:

Iteration 2 : K= 3
Centroids 1 Group 1
$C_1 = \left(\frac{4+7+7+7+7+7}{6}\right) = 6.50 \left(\frac{3+3+3+2+3+3}{6}\right) = 2.83 \quad \left(\frac{19+21+14+19+20+19}{6}\right) = 18.67$
C_1 = (6.50, 2.83, 18.67)

Centroid 2 Group 2
$C_2 = \left(\frac{4+4+3+4+4+7+3+3+3+3}{10}\right) = 3.80$

$\left(\frac{6+6+6+6+9+8+6+6+6+6}{10}\right) = 6.50$

$\left(\frac{14+14+14+14+19+21+21+19+19+19}{10}\right) = 17.40$
C_2 = (3.80, 6.50, 17.40)

Centroid 3 Group 3
$C_3 = \left(\frac{7+4+2+7}{4}\right) = 5 \left(\frac{2+9+6+9}{4}\right) = 6.50 \left(\frac{5+11+5+5}{4}\right) = 6.50$
C_3 = (5, 6.50, 6.50)

So K=3 *Centroids*
C_1 = (6.50, 2.83, 18.67)
C_2 = (3.80, 6.50, 17.40)
C_3 = (5, 6.50, 6.50)

1.A (4,6,14)
C_1 = (6.50, 2.83, 18.67) = $\sqrt{(4-6{,}50)^2 + (6-2{,}83)^2 + (14-18{,}67)^2} = 6.17$
C_2 = (3.80, 6.50, 17.40) = $\sqrt{(4-3{,}80)^2 + (6-6{,}50)^2 + (14-17{,}40)^2} = 3.44$
C_3 = (5, 6.50, 6.50) = $\sqrt{(4-5)^2 + (6-6{,}50)^2 + (14-6{,}50)^2} = 7.58$
2. B (4,6,14)
C_1 = (6.50, 2.83, 18.67) = $\sqrt{(4-6{,}50)^2 + (6-2{,}83)^2 + (14-18{,}67)^2} = 6.17$
C_2 = (3.80, 6.50, 17.40) = $\sqrt{(4-3{,}80)^2 + (6-6{,}50)^2 + (14-17{,}40)^2} = 3.44$
C_3 = (5, 6.50, 6.50) = $\sqrt{(4-5)^2 + (6-6{,}50)^2 + (14-6{,}50)^2} = 7.58$
3. C(7,2,5)
C1 = (6.50, 2.83, 18.67) = $\sqrt{(7-6{,}50)^2 + (2-2{,}83)^2 + (5-18{,}67)^2} = 13.70$
C2 = (3.80, 6.50, 17.40) = $\sqrt{(7-3{,}80)^2 + (2-6{,}50)^2 + (5-17{,}40)^2} = 13.57$
C3 = (5, 6.50, 6.50) = $\sqrt{(7-5)^2 + (2-6{,}50)^2 + (5-6{,}50)^2} = 5.15$
4. D (3,6,14)
C1 = (6.50, 2.83, 18.67) = $\sqrt{(3-6{,}50)^2 + (6-2{,}83)^2 + (14-18{,}67)^2} = 6.64$
C2 = (3.80, 6.50, 17.40) = $\sqrt{(3-3{,}80)^2 + (6-6{,}50)^2 + (14-17{,}40)^2} = 3.53$
C3 = (5, 6.50, 6.50) = $\sqrt{(3-5)^2 + (6-6{,}50)^2 + (14-6{,}50)^2} = 7.78$
5.E (4,6,14)
C1 = (6.50, 2.83, 18.67) = $\sqrt{(4-6{,}50)^2 + (6-2{,}83)^2 + (14-18{,}67)^2} = 6.17$
C2 = (3.80, 6.50, 17.40) = $\sqrt{(4-3{,}80)^2 + (6-6{,}50)^2 + (14-17{,}40)^2} = 3.44$
C3 = (5, 6.50, 6.50) = $\sqrt{(4-5)^2 + (6-6{,}50)^2 + (14-6{,}50)^2} = 7.58$

From the calculation above, the results of the iteration 2 calculation are obtained, which are as shown in the table below.

Table 6: Results of Iteration 2

No	object	X	Y	Z	C_1	C_2	C_3	Group
1	A	4	6	14	6,17	3,44	7.58	2
2	B	4	6	14	6,17	3,44	7.58	2
3	C	7	2	5	13.70	13.57	5,15	3
4	D	3	6	14	6,64	3.53	7,78	2
5	E	4	6	14	6,17	3,44	7.58	2
6	F	4	9	11	10,15	6,87	5,24	3
7	G	4	3	19	2.53	3.85	13.02	1
8	H	7	3	21	2.39	5.95	15.05	1
9	I	4	9	19	6,66	2.97	12.79	2
10	J	2	6	5	14.73	12.54	3.39	3
11	K	7	8	21	5.69	5.04	14.71	2
12	L	7	9	5	15	13.05	3.54	3
13	M	7	3	14	4.70	5.84	8.51	1
14	N	3	6	21	5,27	3.72	14.65	2
15	O	3	6	19	4.73	1.86	12.67	2
16	P	3	6	19	4.73	1.86	12.67	2
17	Q	7	2	19	1.03	5.75	13,44	1

No	object	X	Y	Z	C_1	C_2	C_3	Group
18	R	7	3	20	1.43	5,41	14.09	1
19	S	7	3	19	0.62	5	13,13	1
20	Q	3	6	19	4.73	1.86	12.67	2

After calculating using the *cluster formula* in iteration 2, the groups based on the minimum distance to the nearest *centroid* are:

Old Groups: (2 2 3 2 2 3 1 1 2 3 2 3 1 2 2 2 1 1 1 2)

New Groups : (2 2 3 2 2 3 1 1 2 3 2 3 1 2 2 2 1 1 1 2)

After calculating using the existing *cluster formula, in iteration 2 it is the same as in iteration 1 and there is no data that moves groups again so that the calculation can be stopped. So that a cluster* graph can be made for grouping the number of library members to determine the location of socialization.

3.1. *Clustering* Graph

The following is a *cluster graph* based on the calculation of the results of data mining iterations grouping the number of library members to determine the location of socialization. The graphs obtained are as follows:

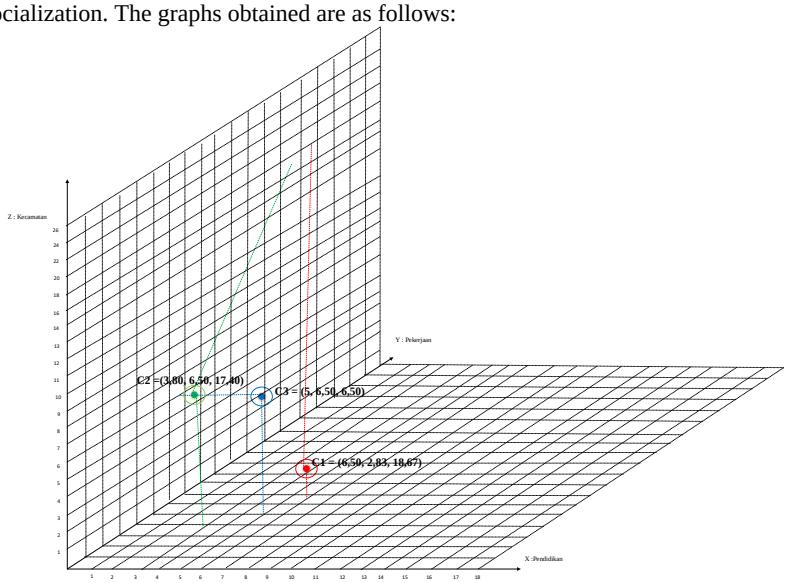


Figure 1: *Cluster* Graph

- *Cluster* 1 : 6,50, 2,83, 18,67 (6, 3, 19)
- *Cluster* 2 : 3,80, 6,50, 17,40 (4, 6, 17)
- *Cluster* 3 : 5, 6,50, 6,50 (5, 6, 6)

Graphic Explanation :

From the 20 data grouping the number of library members to determine the location of socialization at the Langkat Regency Archives and Library service, 3 *groups were obtained , cluster* 1 contained 6 data, *cluster* 2 contained 10 data, and *cluster* 3 contained 4 data.

1. Clusters 1

It can be seen that in *cluster* 1, the cluster results are centered on 6, 3, 1 9, namely D3 education, teacher and sub-district work, namely Stabat with a total of 6 data.

2. Clusters 2

It can be seen that in *cluster* 2, the cluster results are centered on 4, 6, 17, namely high school education, private and sub-district employee jobs , namely completed with a total of 10 data.

3. *Clusters* 3

It can be seen in *cluster* 3, the cluster results are centered on 5, 6, 6, namely D1 education, private and sub-district employee jobs, namely West Brandan with a total of 4 data.

From the results of grouping 20 data above the number of data members , the smallest member is in cluster 3, namely 4 data, thus the location of the socialization needs to be carried out in the West Brandan District area .

3.2. Results Overview

The result description is a data mining *interface* design for grouping the number of library members to determine the location of socialization in which can be described in the menu structure.

1.Home

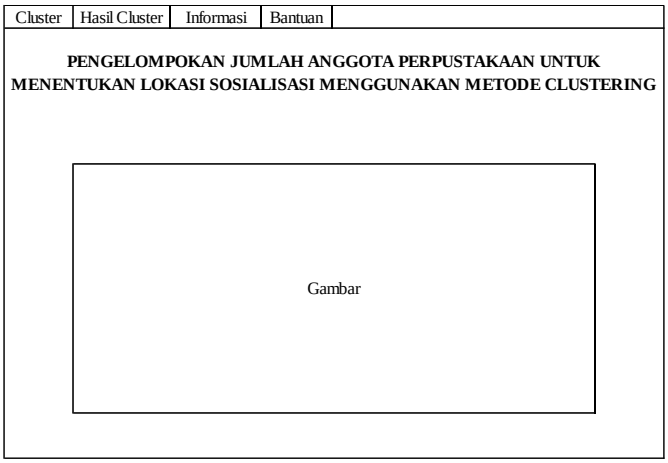


Figure 2: Home _

2. Clusters

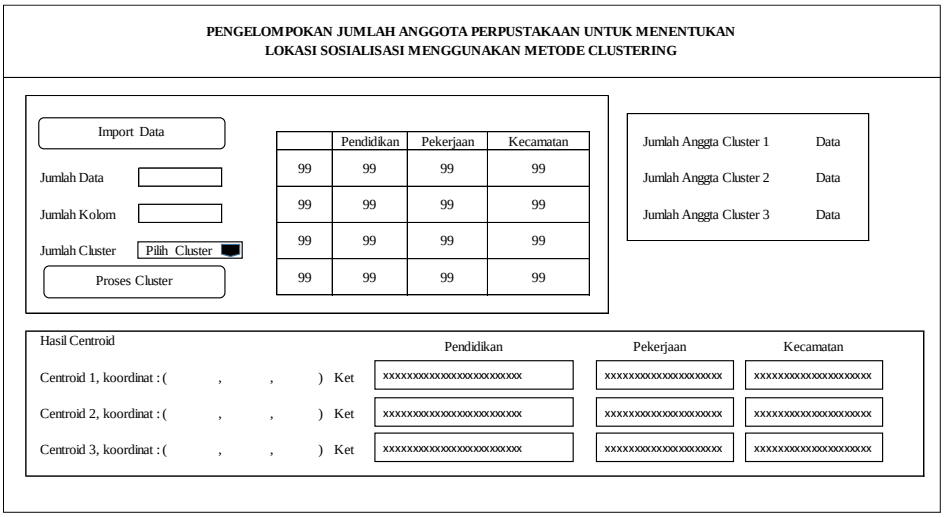
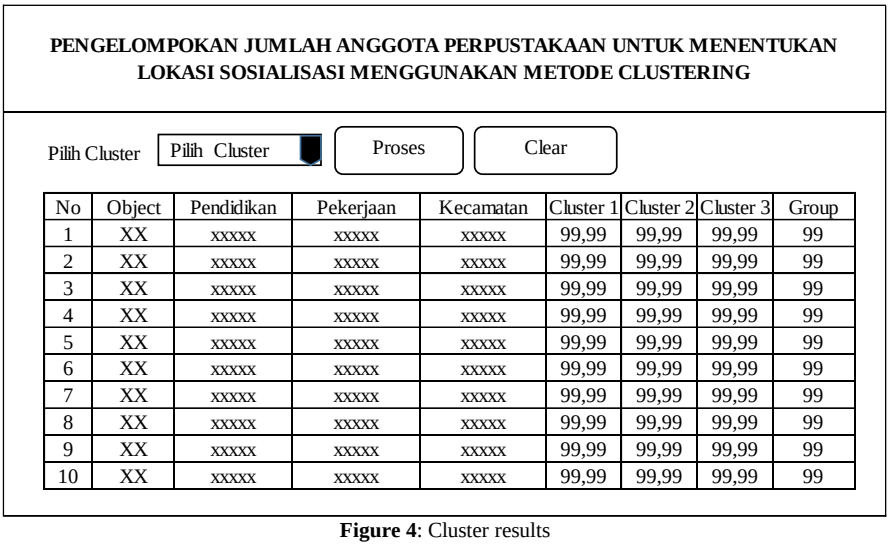


Figure 3: Clusters

3. Cluster Results



No	Object	Pendidikan	Pekerjaan	Kecamatan	Cluster 1	Cluster 2	Cluster 3	Group
1	XX	xxxxx	xxxxx	xxxxx	99,99	99,99	99,99	99
2	XX	xxxxx	xxxxx	xxxxx	99,99	99,99	99,99	99
3	XX	xxxxx	xxxxx	xxxxx	99,99	99,99	99,99	99
4	XX	xxxxx	xxxxx	xxxxx	99,99	99,99	99,99	99
5	XX	xxxxx	xxxxx	xxxxx	99,99	99,99	99,99	99
6	XX	xxxxx	xxxxx	xxxxx	99,99	99,99	99,99	99
7	XX	xxxxx	xxxxx	xxxxx	99,99	99,99	99,99	99
8	XX	xxxxx	xxxxx	xxxxx	99,99	99,99	99,99	99
9	XX	xxxxx	xxxxx	xxxxx	99,99	99,99	99,99	99
10	XX	xxxxx	xxxxx	xxxxx	99,99	99,99	99,99	99

Figure 4: Cluster results

4. Data Information Page

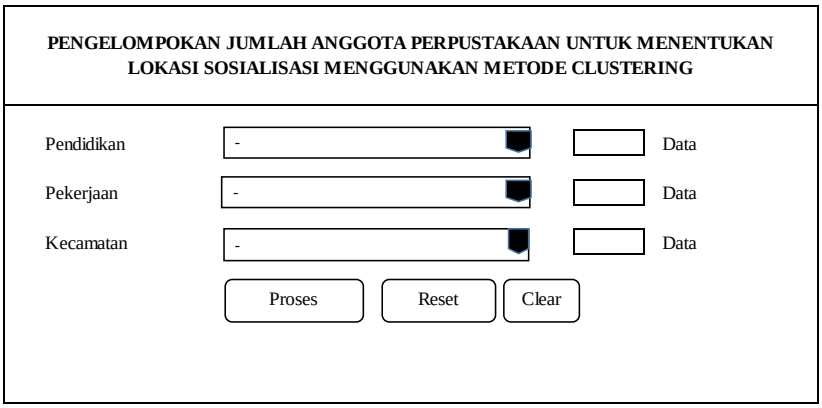


Figure 4 :Data Information

5. Help

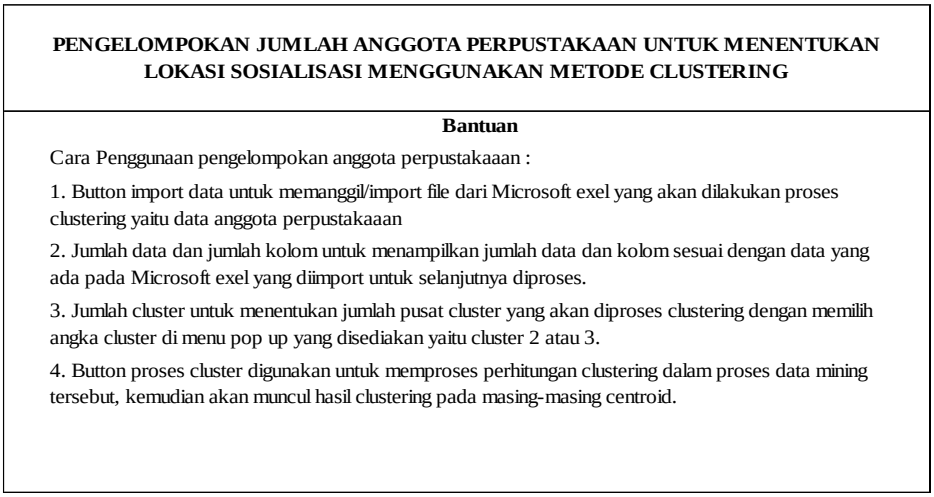


Figure 5: Help

5. **Conclusion**

With the existence of a grouping system for the number of library members to determine the location of this socialization, several conclusions can be drawn, including the following:

1. Helping and facilitating the Langkat Regency Archives and Libraries Office to obtain alternatives and material for consideration in determining the location of the socialization.
2. Providing new information to the Langkat Regency Archives and Library Service regarding people who have an interest in reading in the library.
3. This system has the potential to increase the number of students' interest in reading from elementary to high school, university students and the general public.

Reference

[1] socs.binus.ac.id. (n.d.). https://socs.binus.ac.id/2017/03/09/clustering/.
[2] Adelia Ramadani, Achmad Fauzi, US (2022). Data Mining Grouping Electricity Payments Using the Clustering Method: A Case Study at the Binjai Post Office. El-Mujtama: Journal of Community Service, 2(3), 398–411. https://doi.org/10.47467/elmujtama.v2i3.1947
[3] Budiman, R., & Anto, R. (2019). Application of Data Mining to Determine New Student Admission Promotion Locations at Banten Jaya University (K-Means Clustering Method). ProTekInfo(Information Engineering Research and Observation Development), 6(1), 6. https://doi.org/10.30656/protekinfo.v6i1.1691
[4] Kaisi, EF, & Prasetio, RT (2021). Libraries in DKI Jakarta Using the K-Means Method. Information Systems Proceedings, 2(1), 158–163.
[5] Amril Mutoi Siregar, AP (2016). DATA MINING: Processing Data Into Information with RapidMiner. CV Kekata Group, Suko Harjo.
[6] Cahyo Prianto, SB (2020). Making network disturbance clustering applications using the K-Means clustering method. Nusantara Creative Industry, Bandung.
[7] Nurhayati. (2022b). Ensemble Learning Techniques to Increase Prediction Model Accuracy Performance (Selection of Scholarship Recipient Students). Pascal Books, Tangerang.