

Classification Of Population Data On Status In The Family Based On Last Education And Work Using The Clustering Method (Case Study: Sei Prison Village Office)

Fauziah Ningsih¹, Yani Maulita², Marto Sihombing³

^{1,2,3}STMIK Kaputama

Jl. Veteran No. 4A-9A, Binjai, Sumatera Utara, Indonesia

fauziah21082001@gmail.com^{1*}, yanimaulita26@gmail.com², martosihombing45@gmail.com³

Abstract

Population data is structured individual or individual data through population registration, civil registration and population census activities. It is important to know population data because in making policies and planning regional or state development, population data is needed to describe the condition of an area. Population data include births, deaths, transfers or migration, population composition, population density and so on. This grouping is done so that population data that is already in the archives will be input into an application that will be designed to make it easier for parties who need data without having to look at the data that is still manual. The problems that exist are such as the increase in the number of residents in a city, village or even a district which is increasing while the population that has been recorded still does not have a job, such as status in the family, namely the head of the family is still there who does not work in terms of recent education can still be considered to get a job that matches the last type of education. From the research process conducted on 20 data, 3 groups were obtained, Cluster 1 contained 16 data, Cluster 2 contained 1 data, and Cluster 3 contained 3 data. And the most group obtained is cluster 1, there is education last high school, has a type of work that has not worked and status in the family of the head of the family.

Keywords: data mining, k-means algorithm, population data

1. Introduction

Population data is structured individual or individual data through population registration, civil registration and population census activities. It is important to know population data because in making policies and planning regional or state development, population data is needed to describe the condition of an area. Population data include births, deaths, transfers or migration, population composition, population density and so on.

This grouping is done so that population data that is already in the archives will be input into an application that will be designed to make it easier for parties who need data without having to look at the data that is still manual. The problems that exist are such as the increase in the number of residents in a city, village or even a district which is increasing while the population that has been recorded still does not have a job, such as status in the family, namely the head of the family is still there who does not work in terms of recent education can still be considered to get a job that matches the last type of education. Population growth will also affect economic growth, food prices are uncertain. The processing of the population in this study is grouped based on recent education, employment and status in the family, for example, to find the most clusters, whether there are residents who have higher education but are still not working or vice versa, have low education but are already working, if the data shows that many people are unemployed or there is no work, one of the residents can propose to the Village Head to hold a meeting between residents who do not have a job or who do not have a permanent job to find a solution or form a collaboration between residents with one another in order to create jobs or a joint business. In addition to knowing population data based on recent education and work, it is necessary to design a population data system to find out the identity of population data contained in the data archive at the actual Village Office. Clustering is a data analysis method, which is often included as one of the Data Mining methods, the purpose of which is to group data with the same characteristics and data with different characteristics to one another. Based on the problems above, the writer will influence the criteria (variables) to classify based on the last type of education, type of work and type of status in the family [1], [2].

2. Research Methods

2.1. Data Mining

Data Mining is often also called Knowledge Discovery in Database, is an activity that includes collecting, using historical data to find regularities, patterns or relationships in large data sets. Data Mining is a process that employs one or more computer learning techniques

(machine learning) to analyze and extract knowledge (knowledge) automatically [3]. Data mining is often known as a term used to find hidden knowledge in databases. Data mining is a process that uses statistical, mathematical, artificial intelligence and machine learning techniques to extract and identify useful information and related knowledge from various related databases.

2.2. Definition of Grouping

Grouping or clustering is separating / solving / segmenting data into a number of groups (clusters) according to certain desired characteristics. Grouping can also be interpreted as a process, way, act of grouping. (<https://kbbi.web.id/group>)

2.3. K-Means Algorithm

Clustering is a way to group data that needs to be understood, data mining is part of clustering. That is, it extracts patterns of interest from a large number of data clusters. Clustering is commonly used in business intelligence, image pattern recognition, web search, life sciences, and security. Clustering is the process of grouping data into several clusters so that the data in the clusters have the greatest similarity. Clustering is a method of grouping data that is used to identify groups (clusters) resulting from grouping smaller elements based on similarities to one another.

The 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 a method of non-hierarchical grouping (blocks) that seeks to partition data into clusters/groups so that data that has the same characteristics will be included in the same cluster and data that has different characteristics are grouped into other groups [4].

3. Results And Discussion

3.1. Research Methods

There are several stages of research methodology carried out in solving the problem. These stages are as follows:

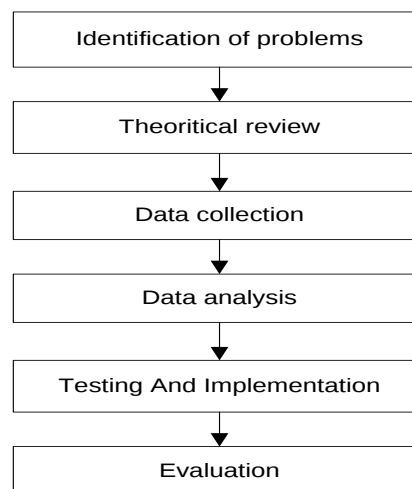


Figure 1: Research Workflow

Based on the picture above, it can be seen that there are several stages in completing research, namely [5]:

1. Problem Identification, this stage is the initial stage in the research, namely by determining the background of the problem, objectives and benefits so as not to get out of focus of discussion or preparation of the thesis.
2. Theory Study, this stage is to look for information, sources related to the problems faced both from literature, journals and the internet as a support and basic basis for thesis writing.
3. Data Collection, this stage is the collection of data needed in making thesis such as interviews, observations and questionnaire methods which can then be processed to the next stage. The interview is to get the right information from a trusted source, the interview is carried out by submitting a number of questions to the source. Observation is an activity towards a data object

process with the intention of being easy to understand and feel. The questionnaire method is a list of written questions that have been prepared beforehand.

4. Data analysis, this stage is the stage of processing and analyzing the data that has been obtained so that the data can be grouped according to the variables that have been determined.
5. Testing and Implementation, this stage is the stage that tests the validation and implementation of the data that has been previously analyzed and the preparation of the program.
6. Evaluation, this stage is the stage of drawing conclusions and suggestions that can be done in the preparation of the thesis. With the conclusion, the results of the entire thesis will be known and it is hoped that with suggestions there will be improvements and benefits for others.

3.2. Flowchart Design

The flowchart of the clustering and K-means hierarchical algorithms is:

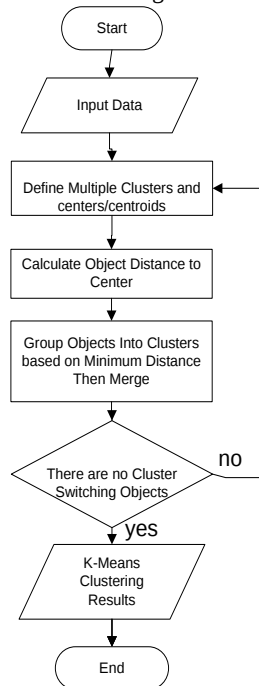


Figure 2: Algorithm Flowchart Hierarchy Clustering and K-means

3.3. Measure Euclidean Distance

In using the clustering method, the initial process for forming clusters is to transform the data into numeric form with predetermined codes, then determine the number of groups (K), calculate the centroids, calculate the object distance to the centroids and then group them based on the distance. closest, if there is an object that moves the group then the iteration is complete [6], [7], [8].

To determine the group of an object, the first thing to do is measure the dEuclidean distance between two object points (X and Y) which is defined as follows:

$$d_{Euclidean}(X,Y) = \sqrt{\sum_i (X_i - Y_i)^2}$$

Existing data is entered into the table for transformation, the required data is 20 data first. The following is the data to be transformed.

Table 1: Data to be Clustered

No	Name	Last education	Work	Status in family
1	Edi Rahman	SLTA	Belum Bekerja	Kepala Keluarga
2	Nur Aidah	SLTP	Ibu Rumah Tangga	Istri
3	Hamad Wahyu	SLTA	Belum Bekerja	Kepala Keluarga
4	Kelik	SLTA	Tidak Mempunyai Pekerjaan Tetap	Kepala Keluarga
5	Siti Zakiah	SLTP	Ibu Rumah Tangga	Istri
6	Liliana	SLTP	Wiraswasta	Istri
7	Candra	SLTP	Tidak Mempunyai Pekerjaan Tetap	Kepala Keluarga
8	Inda Sri Rahayu	SLTA	Wiraswasta	Kepala Keluarga
9	Hamdani	SLTA	Belum Bekerja	Kepala Keluarga
10	Yutika Sari	SLTA	Ibu Rumah Tangga	Istri

11	Suratman	SLTA	Belum Bekerja	Kepala Keluarga
12	Hari Susanti	SLTA	Ibu Rumah Tangga	Istri
13	Seri Agustini	D3	Belum Bekerja	Kepala Keluarga
14	Sellya Hillyza	D3	Tidak Mempunyai Pekerjaan Tetap	Kepala Keluarga
15	Hamdan	SLTA	Tidak Mempunyai Pekerjaan Tetap	Kepala Keluarga
16	Tri Suriati	SLTA	Ibu Rumah Tangga	Istri
17	Pandi Harianto	D3	Wiraswasta	Kepala Keluarga
18	Mhd Syah Putra	SLTP	Ibu Rumah Tangga	Istri
19	Susiani	SLTA	Wiraswasta	Kepala Keluarga
20	Siska Anggrani	SLTP	Ibu Rumah Tangga	Istri

The next step is to transform the data, namely the process of converting or merging data into a suitable format for processing in data mining. Often the data to be used in the data mining process has a format that cannot be used immediately, therefore it is necessary to change the format. From the population data there is the latest education table. The following is the last education table:

Table 2: Last Education

Code	Last Education
1	SD
2	SLTP
3	SLTA
4	D3
5	S1

Table 3: Work

code	Work
1	PNS
2	Wiraswasta
3	Petani
4	Montir
5	Ibu Rumah Tangga
6	Belum Bekerja
7	Tidak Mempunyai Pekerjaan Tetap

Table 4: Status in family

code	Status in family
1	Kepala Keluarga
2	Istri

The next step is to perform data transformation using existing data. The data that has been transformed is obtained from the three existing variables, then transformed according to the code contained in the table for each variable. The following is a data table that has been transformed:

Table 5: Transformed Data

No	Name	Last education	Work	Status in family
1	A	3	6	1
2	B	2	5	2
3	C	3	6	1
4	D	3	7	1
5	E	2	5	2
6	F	2	2	2
7	G	2	7	1
8	H	3	2	1
9	I	3	6	1
10	J	3	5	2
11	K	3	6	1
12	L	3	5	2
13	M	4	6	1
14	N	4	7	1
15	O	3	7	1
16	P	3	5	2
17	Q	4	2	1
18	R	2	5	2
19	S	3	2	1
20	T	2	5	2

Cluster into 3 groups (K=3) and determine the centroid center point.

The clustering calculation process is as follows.

K=3 Centroids

C1 = (3,6,1) taken from data A

C2 = (2,2,2) taken from data F

C3 = (3,2,1) taken from data H

Iterasi 1 :

1. A (3,6,1)

$$C_1 (3,6,1) = \sqrt{(3-3)^2 + (6-6)^2 + (1-1)^2} = 0,00$$

$$C_2 (2,2,2) = \sqrt{(3-2)^2 + (6-2)^2 + (1-2)^2} = 4,24$$

$$C_3 (3,2,1) = \sqrt{(3-3)^2 + (6-2)^2 + (1-3)^2} = 4,00$$
2. B (2,5,2)

$$C_1 (3,6,1) = \sqrt{(2-3)^2 + (5-6)^2 + (2-1)^2} = 1,73$$

$$C_2 (2,2,2) = \sqrt{(2-2)^2 + (5-2)^2 + (2-2)^2} = 3,00$$

$$C_3 (3,2,1) = \sqrt{(2-3)^2 + (5-2)^2 + (2-3)^2} = 3,32$$
3. C (3,6,1)

$$C_1 (3,6,1) = \sqrt{(3-3)^2 + (6-6)^2 + (1-1)^2} = 0,00$$

$$C_2 (2,2,2) = \sqrt{(3-2)^2 + (6-2)^2 + (1-2)^2} = 4,24$$

$$C_3 (3,2,1) = \sqrt{(3-3)^2 + (6-2)^2 + (1-3)^2} = 4,00$$
4. D (3,7,1)

$$C_1 (3,6,1) = \sqrt{(3-3)^2 + (7-6)^2 + (1-1)^2} = 1,00$$

$$C_2 (2,2,2) = \sqrt{(3-2)^2 + (7-2)^2 + (1-2)^2} = 5,20$$

$$C_3 (3,2,1) = \sqrt{(3-3)^2 + (7-2)^2 + (1-3)^2} = 5,00$$
5. E (2,5,2)

$$C_1 (3,6,1) = \sqrt{(2-3)^2 + (5-6)^2 + (2-1)^2} = 1,73$$

$$C_2 (2,2,2) = \sqrt{(2-2)^2 + (5-2)^2 + (2-2)^2} = 3,00$$

$$C_3 (3,2,1) = \sqrt{(2-3)^2 + (5-2)^2 + (2-3)^2} = 3,32$$
6. F (2,2,2)

$$C_1 (3,6,1) = \sqrt{(2-3)^2 + (2-6)^2 + (2-1)^2} = 4,24$$

$$C_2 (2,2,2) = \sqrt{(2-2)^2 + (2-2)^2 + (2-2)^2} = 0,00$$

$$C_3 (3,2,1) = \sqrt{(2-3)^2 + (2-2)^2 + (2-3)^2} = 1,41$$
7. G (2,7,1)

$$C_1 (3,6,1) = \sqrt{(2-3)^2 + (7-6)^2 + (1-1)^2} = 1,41$$

$$C_2 (2,2,2) = \sqrt{(2-2)^2 + (7-2)^2 + (1-2)^2} = 5,10$$

$$C_3 (3,2,1) = \sqrt{(2-3)^2 + (7-2)^2 + (1-3)^2} = 5,10$$
8. H (3,2,1)

$$C_1 (3,6,1) = \sqrt{(3-3)^2 + (2-6)^2 + (1-1)^2} = 4,00$$

$$C_2 (2,2,2) = \sqrt{(3-2)^2 + (2-2)^2 + (1-2)^2} = 1,41$$

$$C_3 (3,2,1) = \sqrt{(3-3)^2 + (2-2)^2 + (1-3)^2} = 0,00$$
9. I (3,6,1)

$$C_1 (3,6,1) = \sqrt{(3-3)^2 + (6-6)^2 + (1-1)^2} = 0,00$$

$$C_2 (2,2,2) = \sqrt{(3-2)^2 + (6-2)^2 + (1-2)^2} = 4,24$$

$$C_3 (3,2,1) = \sqrt{(3-3)^2 + (6-2)^2 + (1-3)^2} = 4,00$$
10. J (3,5,2)

$$C_1 (3,6,1) = \sqrt{(3-3)^2 + (5-6)^2 + (2-1)^2} = 1,41$$

$$C_2 (2,2,2) = \sqrt{(3-2)^2 + (5-2)^2 + (2-2)^2} = 3,16$$

$$C_3 (3,2,1) = \sqrt{(3-3)^2 + (5-2)^2 + (2-3)^2} = 3,16$$
11. K (3,6,1)

$$C_1 (3,6,1) = \sqrt{(3-3)^2 + (6-6)^2 + (1-1)^2} = 0,00$$

$$C_2 (2,2,2) = \sqrt{(3-2)^2 + (6-2)^2 + (1-2)^2} = 4,24$$

$$C_3 (3,2,1) = \sqrt{(3-3)^2 + (6-2)^2 + (1-3)^2} = 4,00$$
12. L (3,5,2)

$$C_1 (3,6,1) = \sqrt{(3-3)^2 + (5-6)^2 + (2-1)^2} = 1,41$$

$$C_2 (2,2,2) = \sqrt{(3-2)^2 + (5-2)^2 + (2-2)^2} = 3,16$$

$$C_3 (3,2,1) = \sqrt{(3-3)^2 + (5-2)^2 + (2-3)^2} = 3,16$$
13. M (4,6,1)

$$C_1 (3,6,1) = \sqrt{(4-3)^2 + (6-6)^2 + (1-1)^2} = 1,00$$

$$C_2 (2,2,2) = \sqrt{(4-2)^2 + (6-2)^2 + (1-2)^2} = 4,58$$

$$C_3 (3,2,1) = \sqrt{(4-3)^2 + (6-2)^2 + (1-3)^2} = 4,12$$
14. N (4,7,1)

$$C_1 (3,6,1) = \sqrt{(4-3)^2 + (7-6)^2 + (1-1)^2} = 1,41$$

$$C_2 (2,2,2) = \sqrt{(4-2)^2 + (7-2)^2 + (1-2)^2} = 5,48$$

$$C_3 (3,2,1) = \sqrt{(4-3)^2 + (7-2)^2 + (1-3)^2} = 5,10$$

15. O (3,7,1)
 $C_1 (3,6,1) = \sqrt{(3-3)^2 + (7-6)^2 + (1-1)^2} = 1,00$
 $C_2 (2,2,2) = \sqrt{(3-2)^2 + (7-2)^2 + (1-2)^2} = 5,20$
 $C_3 (3,2,1) = \sqrt{(3-3)^2 + (7-2)^2 + (1-3)^2} = 5,00$
16. P (3,5,2)
 $C_1 (3,6,1) = \sqrt{(3-3)^2 + (5-6)^2 + (2-1)^2} = 1,41$
 $C_2 (2,2,2) = \sqrt{(3-2)^2 + (5-2)^2 + (2-2)^2} = 3,16$
 $C_3 (3,2,1) = \sqrt{(3-3)^2 + (5-2)^2 + (2-3)^2} = 3,16$
17. Q (4,2,1)
 $C_1 (3,6,1) = \sqrt{(4-3)^2 + (2-6)^2 + (1-1)^2} = 1,41$
 $C_2 (2,2,2) = \sqrt{(4-2)^2 + (2-2)^2 + (1-2)^2} = 3,16$
 $C_3 (3,2,1) = \sqrt{(4-3)^2 + (2-2)^2 + (1-3)^2} = 3,16$
18. R (2,5,2)
 $C_1 (3,6,1) = \sqrt{(2-3)^2 + (5-6)^2 + (2-1)^2} = 1,73$
 $C_2 (2,2,2) = \sqrt{(2-2)^2 + (5-2)^2 + (2-2)^2} = 3,00$
 $C_3 (3,2,1) = \sqrt{(2-3)^2 + (5-2)^2 + (2-3)^2} = 3,32$
19. S (3,2,1)
 $C_1 (3,6,1) = \sqrt{(3-3)^2 + (2-6)^2 + (1-1)^2} = 4,00$
 $C_2 (2,2,2) = \sqrt{(3-2)^2 + (2-2)^2 + (1-2)^2} = 1,41$
 $C_3 (3,2,1) = \sqrt{(3-3)^2 + (2-2)^2 + (1-3)^2} = 0,00$
20. T (2,5,2)
 $C_1 (3,6,1) = \sqrt{(2-3)^2 + (5-6)^2 + (2-1)^2} = 1,73$
 $C_2 (2,2,2) = \sqrt{(2-2)^2 + (5-2)^2 + (2-2)^2} = 3,00$
 $C_3 (3,2,1) = \sqrt{(2-3)^2 + (5-2)^2 + (2-3)^2} = 3,32$

The results of iteration 1 calculations are obtained from the table that has been transformed and then the calculations are carried out to get the results of iteration 1. After the results of iteration 1 are obtained, then they are entered into the iteration table 1. The following is a table of the results of iteration 1:

Table 6: Iteration Calculation Results 1

No	X	Y	Z	C1	C2	C3	Group
1	3	6	1	0,00	4,24	4,00	1
2	2	5	2	1,73	3,00	3,32	1
3	3	6	1	0,00	4,24	4,00	1
4	3	7	1	1,00	5,20	5,00	1
5	2	5	2	1,73	3,00	3,32	1
6	2	2	2	4,24	0,00	1,41	2
7	2	7	1	1,41	5,10	5,10	1
8	3	2	1	4,00	1,41	0,00	3
9	3	6	1	0,00	4,24	4,00	1
10	3	5	2	1,41	3,16	3,16	1
11	3	6	1	0,00	4,24	4,00	1
12	3	5	2	1,41	3,16	3,16	1
13	4	6	1	1,00	4,58	4,12	1
14	4	7	1	1,41	5,48	5,10	1
15	3	7	1	1,00	5,20	5,00	1
16	3	5	2	1,41	3,16	3,16	1
17	4	2	1	4,12	2,24	1,00	3
18	2	5	2	1,73	3,00	3,32	1
19	3	2	1	4,00	1,41	0,00	3
20	2	5	2	1,73	3,00	3,32	1

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

Old Group : {0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}

New Group : {1,1,1,1,1,2,1,2,1,1,1,1,1,1,1,3,1,3,1}

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

Iteration 2 :

K = 3

Centroids 1 Group 1

$C_1 = \frac{(3+2+3+3+2+2+3+3+3+3+4+4+3+3+2+2)}{16}$

16

$\frac{(6+5+6+7+5+7+6+5+6+5+6+7+5+5+5)}{16}$

16

$$\frac{(1+2+1+1+2+1+1+2+1+1+1+2+2+2))}{16}$$

$$= (2,81 \ 5,81 \ 1,44)$$

Centroid 2 Group 2

$$C_2 = \frac{\begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}}{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}$$

$$= (2,00 \ 2,00 \ 2,00)$$

Centroid 3 Group 3

$$C_3 = \frac{\begin{pmatrix} 3+4+3 \\ 2+2+2 \\ 1+1+1 \end{pmatrix}}{\begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}}$$

$$= (3,33 \ 2,00 \ 1,00)$$

So K=3 Centroid

$$C_1 = (2,81 \ 5,81 \ 1,44)$$

$$C_2 = (2,00 \ 2,00 \ 2,00)$$

$$C_3 = (3,33 \ 2,00 \ 1,00)$$

1. A(3,6,1)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(3 - 2,81)^2 + (6 - 5,81)^2 + (1 - 1,44)^2} = 0,51$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(3 - 2,00)^2 + (6 - 2,00)^2 + (1 - 2,00)^2} = 4,24$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(3 - 3,33)^2 + (6 - 2,00)^2 + (1 - 1,00)^2} = 4,10$$

2. B(2,5,2)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(2 - 2,81)^2 + (5 - 5,81)^2 + (2 - 1,44)^2} = 1,28$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(2 - 2,00)^2 + (5 - 2,00)^2 + (2 - 2,00)^2} = 3,00$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(2 - 3,33)^2 + (5 - 2,00)^2 + (2 - 1,00)^2} = 3,43$$

3. C(3,6,1)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(3 - 2,81)^2 + (6 - 5,81)^2 + (1 - 1,44)^2} = 0,51$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(3 - 2,00)^2 + (6 - 2,00)^2 + (1 - 2,00)^2} = 4,24$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(3 - 3,33)^2 + (6 - 2,00)^2 + (1 - 1,00)^2} = 4,10$$

4. D(3,7,1)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(3 - 2,81)^2 + (7 - 5,81)^2 + (1 - 1,44)^2} = 1,28$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(3 - 2,00)^2 + (7 - 2,00)^2 + (1 - 2,00)^2} = 5,20$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(3 - 3,33)^2 + (7 - 2,00)^2 + (1 - 1,00)^2} = 5,01$$

5. E(2,5,2)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(2 - 2,81)^2 + (5 - 5,81)^2 + (2 - 1,44)^2} = 1,28$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(2 - 2,00)^2 + (5 - 2,00)^2 + (2 - 2,00)^2} = 3,00$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(2 - 3,33)^2 + (5 - 2,00)^2 + (2 - 1,00)^2} = 3,43$$

6. F(2,2,2)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(2 - 2,81)^2 + (2 - 5,81)^2 + (2 - 1,44)^2} = 3,94$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(2 - 2,00)^2 + (2 - 2,00)^2 + (2 - 2,00)^2} = 0,00$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(2 - 3,33)^2 + (2 - 2,00)^2 + (2 - 1,00)^2} = 1,67$$

7. G(2,7,1)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(2 - 2,81)^2 + (7 - 5,81)^2 + (1 - 1,44)^2} = 1,50$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(2 - 2,00)^2 + (7 - 2,00)^2 + (1 - 2,00)^2} = 5,10$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(2 - 3,33)^2 + (7 - 2,00)^2 + (1 - 1,00)^2} = 5,17$$

8. H(3,2,1)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(3 - 2,81)^2 + (2 - 5,81)^2 + (1 - 1,44)^2} = 3,84$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(3 - 2,00)^2 + (2 - 2,00)^2 + (1 - 2,00)^2} = 1,41$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(3 - 3,33)^2 + (2 - 2,00)^2 + (1 - 1,00)^2} = 0,33$$

9. I (3,6,1)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(3 - 2,81)^2 + (6 - 5,81)^2 + (1 - 1,44)^2} = 0,51$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(3 - 2,00)^2 + (6 - 2,00)^2 + (1 - 2,00)^2} = 4,24$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(3 - 3,33)^2 + (6 - 2,00)^2 + (1 - 1,00)^2} = 4,10$$

10. J (3,5,2)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(3 - 2,81)^2 + (5 - 5,81)^2 + (2 - 1,44)^2} = 1,01$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(3 - 2,00)^2 + (5 - 2,00)^2 + (2 - 2,00)^2} = 3,16$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(3 - 3,33)^2 + (5 - 2,00)^2 + (2 - 1,00)^2} = 3,18$$

11. K (3,6,1)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(3 - 2,81)^2 + (6 - 5,81)^2 + (1 - 1,44)^2} = 0,51$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(3 - 2,00)^2 + (6 - 2,00)^2 + (1 - 2,00)^2} = 4,24$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(3 - 3,33)^2 + (6 - 2,00)^2 + (1 - 1,00)^2} = 4,10$$

12. L (3,5,2)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(3 - 2,81)^2 + (5 - 5,81)^2 + (2 - 1,44)^2} = 1,01$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(3 - 2,00)^2 + (5 - 2,00)^2 + (2 - 2,00)^2} = 3,16$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(3 - 3,33)^2 + (5 - 2,00)^2 + (2 - 1,00)^2} = 3,18$$

13. M (4,6,1)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(4 - 2,81)^2 + (6 - 5,81)^2 + (1 - 1,44)^2} = 1,28$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(4 - 2,00)^2 + (6 - 2,00)^2 + (1 - 2,00)^2} = 4,58$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(4 - 3,33)^2 + (6 - 2,00)^2 + (1 - 1,00)^2} = 4,06$$

14. N (4,7,1)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(4 - 2,81)^2 + (7 - 5,81)^2 + (1 - 1,44)^2} = 1,74$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(4 - 2,00)^2 + (7 - 2,00)^2 + (1 - 2,00)^2} = 5,48$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(4 - 3,33)^2 + (7 - 2,00)^2 + (1 - 1,00)^2} = 5,04$$

15. O (3,7,1)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(3 - 2,81)^2 + (7 - 5,81)^2 + (1 - 1,44)^2} = 1,28$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(3 - 2,00)^2 + (7 - 2,00)^2 + (1 - 2,00)^2} = 5,20$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(3 - 3,33)^2 + (7 - 2,00)^2 + (1 - 1,00)^2} = 5,01$$

16. P (3,5,2)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(3 - 2,81)^2 + (5 - 5,81)^2 + (2 - 1,44)^2} = 1,01$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(3 - 2,00)^2 + (5 - 2,00)^2 + (2 - 2,00)^2} = 3,16$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(3 - 3,33)^2 + (5 - 2,00)^2 + (2 - 1,00)^2} = 3,18$$

17. Q (4,2,1)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(4 - 2,81)^2 + (2 - 5,81)^2 + (1 - 1,44)^2} = 4,02$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(4 - 2,00)^2 + (2 - 2,00)^2 + (1 - 2,00)^2} = 2,24$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(4 - 3,33)^2 + (2 - 2,00)^2 + (1 - 1,00)^2} = 0,67$$

18. R (2,5,2)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(2 - 2,81)^2 + (5 - 5,81)^2 + (2 - 1,44)^2} = 1,28$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(2 - 2,00)^2 + (5 - 2,00)^2 + (2 - 2,00)^2} = 3,00$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(2 - 3,33)^2 + (5 - 2,00)^2 + (2 - 1,00)^2} = 3,43$$

19. S (3,2,1)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(3 - 2,81)^2 + (2 - 5,81)^2 + (1 - 1,44)^2} = 3,84$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(3 - 2,00)^2 + (2 - 2,00)^2 + (1 - 2,00)^2} = 1,41$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(3 - 3,33)^2 + (2 - 2,00)^2 + (1 - 1,00)^2} = 0,33$$

20. T (2,5,2)

$$C_1(2,81 \ 5,81 \ 1,44) = \sqrt{(2 - 2,81)^2 + (5 - 5,81)^2 + (2 - 1,44)^2} = 1,28$$

$$C_2(2,00 \ 2,00 \ 2,00) = \sqrt{(2 - 2,00)^2 + (5 - 2,00)^2 + (2 - 2,00)^2} = 3,00$$

$$C_3(3,33 \ 2,00 \ 1,00) = \sqrt{(2 - 3,33)^2 + (5 - 2,00)^2 + (2 - 1,00)^2} = 3,43$$

The results of the iteration 2 calculations are obtained from the results of the iteration calculation table 1 and then obtained from the Cluster in iteration 1. The following is a table of the results of iteration 2:

Table 7: Iteration Calculation Results 2

No	X	Y	Z	C1	C2	C3	Group
1	3	6	1	0,51	4,24	4,01	1
2	2	5	2	1,28	3,00	3,43	1
3	3	6	1	0,51	4,24	4,01	1
4	3	7	1	1,28	5,20	5,01	1
5	2	5	2	1,28	3,00	3,43	1
6	2	2	2	3,94	0,00	1,67	2
7	2	7	1	1,50	5,10	5,17	1
8	3	2	1	3,84	1,41	0,33	3
9	3	6	1	0,51	4,24	4,01	1
10	3	5	2	1,01	3,16	3,18	1
11	3	6	1	0,51	4,24	4,01	1
12	3	5	2	1,01	3,16	3,18	1
13	4	6	1	1,28	4,58	4,06	1
14	4	7	1	1,74	5,48	5,04	1
15	3	7	1	1,28	5,20	5,01	1
16	3	5	2	1,01	3,16	3,18	1
17	4	2	1	4,02	2,24	0,67	3
18	2	5	2	1,28	3,00	3,43	1
19	3	2	1	3,84	1,41	0,33	3
20	2	5	2	1,28	3,00	3,43	1

After calculating using the existing cluster formula, iteration 1 occurs the same as iteration 2 and there is no data that moves groups again so the calculation can be stopped. So that the cluster graph can be made. The group results obtained from the calculation of Iterations 1 and 2 are as follows:

Old Group : {1,1,1,1,1,2,2,1,1,1,1,1,1,1,3,1,3,1}

New Group : {1,1,1,1,1,2,2,1,1,1,1,1,1,1,3,1,3,1}

4. Conclusion

From the research process conducted on 20 data, 3 groups were obtained, Cluster 1 contained 16 data, Cluster 2 contained 1 data, and Cluster 3 contained 3 data. And the most group obtained is cluster 1, there is education last high school, has a type of work that has not worked and status in the family of the head of the family.

References

- [1] Arhami, M., & Nasir, M. (2020). Data Mining (R. Indah utami, Ed.; 1st ed.). cv andi offset.
- [2] Buaton, R., Saragih, R., & SistemInformasi STMIK Kaputama Binjai Sumatera Utara, P. (2022). Data Mining Pengelompokan Akta Nikah Berdasarkan Usia Nikah atau Domisili Menggunakan Metode Clustering: Studi Kasus Kemenag Langkat. In Sci-Tech Journal (Vol. 2, Issue 1).
- [3] Dr. Eng. Ags Naba, Pengertian GUIDE atau GUI Yogyakarta, 2008
- [4] Eko Prasetyo, Data Mining Konsep Dan Aplikasi Menggunakan Matlab, PT.Andi, Yogyakarta, 2012
- [5] Irmanita Nasution, Agus Perdana Windarto, M.Fauzan. 2020. "Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi." Vol. 2, No.2, Hal: 76-83 Desember 2020
- [6] Kusri, Emha Taufiq Luthfi, Algoritma Data Mining, Penerbit C.V Andi, Yogyakarta 2009
- [7] Lina Listiani, Yoga Handoko Agustin, Mochammad Zaenal Ramdhani. 2019. "Implementasi Algoritma K-Means Clustering Untuk Rekomendasi Pekerjaan Berdasarkan Pengelompokan Data Penduduk." Seminar Nasional Sistem Informasi dan Teknik Informatika.
- [8] Yatini B, Definisi Flowchart, Jakarta PT. Gramedia Widia Sarana Indonesia, 2006