



Village Status Grouping Analysis Using Agglomerative Hierarchical Clustering (AHC)

Paska Marto Hasugian

Teknik Informatika

STMIK Pelita Nusantara, Medan, Jl. Iskandar Muda No 1. Sumatera Utara, 20154 Indonesia

E-mail: paskamarto86@gmail.com

ARTICLE INFO

Article history:

Received: 20 April 2020

Revised: 22 Mei 2020

Accepted: 28 Mei 2020

Keywords:

Teknologi;

Big Data;

Cluster;

AHC;

Village;

ABSTRACT

Increasing the standard of living of the community, both materially and spiritually. Careful development planning can create a development that is right on target. One of the supporting aspects of development planning is the availability of detailed data at the smallest area level. Information to the smallest area can be used as a guide in making policies that are more targeted. The village as a government that is directly in contact with the community is the main focus in government development, this is because most of Indonesia's territory is in rural areas which are directly under the auspices of the PDPTT Ministry (Ministry of Villages, Disadvantaged Village Development, and Transmigration), to carry out their duties and The responsibility for village development has been assessed in establishing a village classification based on the building village index to convert an underdeveloped village into a developed or independent village which is assessed based on certain sectors through a very long and certainly sustainable process. Determination of classification is based on the village index value build (IDM) based on the average Ecological Resilience Index (IKL), Economic Resilience Index (IKE), and Social Resilience Index (IKS). From the IDM value, the villages will be classified in the status of villages namely independent villages, Advanced, Developing, Underdeveloped, and Very Underdeveloped Villages. In this study, testing was carried out using the AHC algorithm

Copyright © 2020 Jurnal Mantik.

All rights reserved.

1. Introduction

Villages, in general, are population centers whose eyes are mainly on agriculture and are located far from cities. Administratively, a village is an administrative unit headed by a village head. Village is the lowest regional government. This is clarified by Law No. 5 of 1979. In-Law no. 5 of 1979 (Law No.5 / 79) concerning Village Government, what is meant by the village is an area occupied by several residents as a community unit, including a legal community unit that has an organization directly under the Camat and has the right to run its household. Within the ties of the Unitary State of the Republic of Indonesia and geographically, the definition of village is a geographical manifestation caused by physiographic, socio-economic, political, and cultural elements in their relationship and mutual influence with other regions. Village development directly under the auspices of the Ministry of Villages, Development of Disadvantaged Areas, and Transmigration (KDPDPTT) of the Republic of Indonesia is a ministry within the Government of Indonesia which deals with the development of villages and rural areas, empowerment of rural communities, accelerate the development of disadvantaged areas, and transmigration. The Ministry of Villages, Development of Disadvantaged Regions, and Transmigration is under and responsible to the President. This ministry is led by a Minister for Villages. [1]

Development of Disadvantaged Areas, and Transmigras with the main task of referring to Permendes No. 6 of 2015. Of the many tasks and responsibilities of the village ministry, one of which has been running and continuing to date is the classification of villages based on the value of Index village development (IDM). The IDM is a database (baseline) for village development which is the basis for assessing the progress and independence of the village, one of the inputs (focus) in the formulation of strategic issues and main problems related to the development and empowerment of the Village community, input in the formulation of targeting (location targets) related to national development targets and coordination instruments between Ministries / Institutions, local and village governments, for effective achievement of national development targets, the



process of determining IDM by assessing several sectors namely ecological resilience (IKL), Economic Resilience index (IKE) and index Social Resilience (IKS). [2]. The combination of the three sectors is a combination that will determine the quality of the village. The assessment process was carried out to obtain IDM scores by forming a team that immediately jumped into the field by assessing these three aspects, then calculated and adjusted to the predetermined classification range. [3], [4]

The method used to solve village grouping is Agglomerative Hierarchical Clustering, which is a method using a Bottom-Up design strategy that starts by placing each object as a separate cluster (atomic cluster) and then combining the atomic clusters into larger and more clustered clusters [5]–[12]. Large again until finally all objects are united in a cluster or the process can also stop if it has reached a certain condition. The distance measure used to combine two cluster objects is Euclidean Distance, which can be seen in the following equation. [3], [4]

2. Method

The stages of pattern formation are as follows

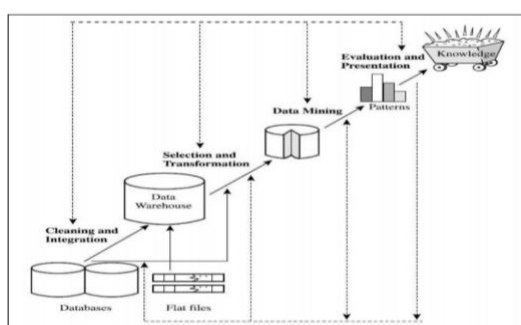


Fig 1. Pattern Forming Process

The explanation from table 1 follows:

- a) Data cleaning, eliminating noise, and inconsistent data.
- b) Data integration, combining data from a variety of different data sources
- c) Data selection, retrieving data relevant to the analysis task from the database
- d) Data transformation, Transforming, or combining data into a form suitable for extracting through summary or aggregation operations.
- e) Data mining, an essential process for extracting patterns from data with intelligent methods.

3. Result and Discussion

The data used in the grouping process is the IDM value with a combination of village status with random data references based on the village ministry, development of disadvantaged areas, and transmigration, with the following test data.

Tabel 1.
Testing Data Samples

NO	IDM	IKS	IKE	IKL
1	0.617	0.6343	0.55	0.6667
2	0.5835	0.6171	0.4667	0.6667
3	0.5184	0.5886	0.5	0.4667
4	0.5789	0.72	0.55	0.4667
5	0.5292	0.5543	0.4333	0.6
6	0.4717	0.5486	0.3333	0.5333
7	0.5894	0.6514	0.45	0.6667
8	0.5214	0.5143	0.3833	0.6667
9	0.5781	0.6343	0.4333	0.6667
10	0.5273	0.5486	0.3667	0.6667
11	0.557	0.5543	0.45	0.6667
12	0.683	0.5657	0.8167	0.6667
13	0.5451	0.4686	0.5	0.6667
14	0.543	0.4457	0.5167	0.6667

The data will be tested by means of the following data transformation:

Table 2.
Data Transformation

NO	IDM	IKS	IKE	IKL
1	0.617	0.6343	0.55	0.6667
2	0.5835	0.6171	0.4667	0.6667
3	0.5184	0.5886	0.5	0.4667
4	0.5789	0.72	0.55	0.4667
5	0.5292	0.5543	0.4333	0.6

Discussion

a) Calculating the distance for all pairs of two data with the following formula

$$D(\text{data1}, \text{data1}) = |0.51-2| + |0.51-2| = 0$$

In the same way for all data, the following table is formed

Table 3.
Distance of two Data Pairs

DMAN	1	2	3	4	5	6	7	8	9	10
1	0	1.06	-2.28	-3.3	-1.08	-1.16	-2.23	1.06	-1.16	-1.16
2	-1.06	0	-3.34	-4.36	-2.14	-2.22	-3.29	0	-2.22	-2.22
3	2.28	3.34	0	-1.02	1.2	1.12	0.05	3.34	1.12	1.12
4	3.3	4.36	1.02	0	2.22	2.14	1.07	4.36	2.14	2.14
5	-0.92	-1.86	0.8	1.78	0	-0.08	0.85	-1.86	-0.08	-0.08
6	1.16	2.22	-1.12	-2.14	0.08	0	-1.07	2.22	0	0
7	-0.06	0	-0.34	-0.36	-0.14	-0.22	-0.29	0	-0.22	-0.22
8	-2	0	-6	-8	-4	-4	-6	0	-4	-4
9	1.16	2.22	-1.12	-2.14	0.08	0	-1.07	2.22	0	0
10	1.16	2.22	-1.12	-2.14	0.08	0	-1.07	2.22	0	0

after the data is presented, the data comparison process is carried out using the single linkage method by selecting the two smallest groups, namely data 8 and data 2 so that the two groups are combined $\text{Min}(D_{man}) = \min(2,8) = 8$.

1. Calculating the distance between groups (2,8) and the remaining group (1,3,4,5,6,7,9,10)

$$D(28) = \min\{21.81\} = \min(-1.06, -2) = -2$$

$$D(28) = \min\{23.83\} = \min(1.22, 0.28) = 0.28$$

$$D(28) = \min\{24.84\} = \min(2.24, 1.3) = 1.5$$

$$D(28) = \min\{25.85\} = \min(-1.98, -2.92) = -2.92$$

$$D(28) = \min\{26.86\} = \min(0.1, -0.84) = -0.84$$

$$D(28) = \min\{27.87\} = \min(-1.12, -2.06) = -2.06$$

$$D(28) = \min\{29.89\} = \min(0.1, -0.84) = -0.84$$

$$D(28) = \min\{210,810\} = \min(0.1, -0.84) = -0.84$$

b) Deleting the corresponding Matrix As well as combining the matrix into one column

Table 4.
Corresponding matrix

DMAN	1	2	3	4	5	6	7	8	9	10
1	0	1.06	-2.28	-3.3	-1.08	-1.16	-2.23	1.06	-1.16	-1.16
2	-1.06	0	-3.34	-4.36	-2.14	-2.22	-3.29	0	-2.22	-2.22
3	2.28	3.34	0	-1.02	1.2	1.12	0.05	3.34	1.12	1.12
4	3.3	4.36	1.02	0	2.22	2.14	1.07	4.36	2.14	2.14
5	-0.92	-1.86	0.8	1.78	0	-0.08	0.85	-1.86	-0.08	-0.08
6	1.16	2.22	-1.12	-2.14	0.08	0	-1.07	2.22	0	0
7	-0.06	0	-0.34	-0.36	-0.14	-0.22	-0.29	0	-0.22	-0.22
8	-2	0	-6	-8	-4	-4	-6	0	-4	-4
9	1.16	2.22	-1.12	-2.14	0.08	0	-1.07	2.22	0	0
10	1.16	2.22	-1.12	-2.14	0.08	0	-1.07	2.22	0	0

From this table, the following new matrix is generated



Table 5.
New Matrices

DMAN	1	2.8	3	4	5	6	7	9	10
1	0	1.06	-2.28	-3.3	-1.08	-1.16	-2.23	-1.16	-1.16
2.8	0	0	0.28	1.5	2.92	-0.84	-2.06	-0.84	-0.84
3	2.28	3.34	0	-1.02	1.2	1.12	0.05	1.12	1.12
4	3.3	4.36	1.02	0	2.22	2.14	1.07	2.14	2.14
5	-0.92	-1.86	0.8	1.78	0	-0.08	0.85	-0.08	-0.08
6	1.16	2.22	-1.12	-2.14	0.08	0	-1.07	0	0
7	-0.06	0	-0.34	-0.36	-0.14	-0.22	-0.29	-0.22	-0.22
9	1.16	2.22	-1.12	-2.14	0.08	0	-1.07	0	0
10	1.16	2.22	-1.12	-2.14	0.08	0	-1.07	0	0

c) Selecting the two smallest groups and forming a new matrix with at this solution the smallest data is the seventh and fourth data (7,4). The adjustment process is carried out with the remaining data combination of 1, 2.8,3,5,6,9,10. With the following calculations:

$$D(7,4) = \min \{7.1,4.1\} = \min (-0.06,3.3) = - 0.06$$

$$D(7,4) = \min \{7.28,4.28\} = \min (-0.06,3.3) = - 0.06$$

$$D(7,4) = \min \{7.3,4.3\} = \min (2.22,5.58) = 2.22$$

$$D(7,4) = \min \{7.5,4.5\} = \min (-0.98,2.38) = - 0.98$$

$$D(7,4) = \min \{7.6,4.6\} = \min (1.1,4.46) = 1.1$$

$$D(7,4) = \min \{7.9,4.9\} = \min (1.1,4.46) = 1.1$$

$$D(7,4) = \min \{7.10,4.10\} = \min (1.1,4.46) 1.1$$

Resulting in the following new table

Table 6.
New Group of Corresponding Matrix

DMAN	1	2.8	3	5	6	7,4	9	10
1	0	1.06	-2.28	-1.08	-1.16	-2.23	-1.16	-1.16
2.8	0	0	0.28	2.92	-0.84	-2.06	-0.84	-0.84
3	2.28	3.34	0	1.2	1.12	0.05	1.12	1.12
5	-0.92	-1.86	0.8	0	-0.08	0.85	-0.08	-0.08
6	1.16	2.22	-1.12	0.08	0	-1.07	0	0
7,4	-0.06	-0.06	2.22	0.98	1.1	0	1.1	1.1
9	1.16	2.22	-1.12	0.08	0	-1.07	0	0
10	1.16	2.22	-1.12	0.08	0	-1.07	0	0

Forms a new group with the following two smallest values, From these data, the smallest data is 7.4 and data 5, so the following calculations are generated:

$$D(7.4, 5) = \min \{7.4,1,5,1\} = \min (-0.06, -0.92) = - 0.92$$

$$D(7.4, 5) = \min \{74,28, 5,28\} = \min (-0.06, -0.92) = - 0.92$$

$$D(7.4, 5) = \min \{74,3, 5,3\} = \min (2.22,1.36) = 1.36$$

$$D(7.4, 5) = \min \{74,6, 5,6\} = \min (1.1,0.24) = 0.24$$

$$D(7.4, 5) = \min \{7,4,9, 5,9\} = \min (1.1,0.24) = 0.24$$

$$D(7.4, 5) = \min \{7,4,10, 5,10\} = \min (1.1,0.24) = 0.24$$

Table 7.
Smallest New Groups

DMAN	1	2.8	3	6	74,5	9	10
1	0	1.06	-2.28	-1.16	-2.23	-1.16	-1.16
2.8	0	0	0.28	-0.84	-2.06	-0.84	-0.84
3	2.28	3.34	0	1.12	0.05	1.12	1.12
7,45	-0.92	-0.92	1.36	0,24	0	0,24	0,24
6	1.16	2.22	-1.12	0	-1.07	0	0
9	1.16	2.22	-1.12	0	-1.07	0	0
10	1.16	2.22	-1.12	0	-1.07	0	0

Followed by forming a new table by combining the two smallest data by calculating the closest distance. The calculation process follows:

$$D(1,745) = \min \{745,28, 1,28\} = \min (-0.92,0) = - 0.92$$



$$D(1,745) = \min \{745.3, 1.3\} = \min (-0.92, 0) = -0.92$$

$$D(1,745) = \min \{745.6, 1.6\} = \min (0.24, 1.16) = 0.24$$

$$D(1,745) = \min \{745.9, 1.9\} = \min (0.24, 1.16) = 0.24$$

$$D(1,745) = \min \{745.10, 1.10\} = \min (0.24, 1.16) = 0.24$$

Table 8.
Combining the Smallest Data

DMAN	1.745	2.8	3	6	9	10
1.745	0	1.06	-2.28	-1.16	-1.16	-1.16
2.8	0	0	0.28	-0.84	-0.84	-0.84
3	2.28	3.34	0	1.12	1.12	1.12
6	1.16	2.22	-1.12	0	0	0
9	1.16	2.22	-1.12	0	0	0
10	1.16	2.22	-1.12	0	0	0

The process continues to form groups for each data that has been presented in this test, several groups are determined with their respective closeness, namely data grouping 1 (1,4,5,7) data 2 (2,8) third data (3) and data 4 (6,9,10).

d) Conclusion

The presentation of clustering result data is presented in iteration form which is displayed in graphical form. And displayed all iterations included in the clustering calculation. The results of this iterative calculation can be used to analyze the clustered data, the application of the AHC clustering method in grouping village status data and to help analyze previous village data. Information obtained from the application of the AHC clustering method can be used to help make decisions to solve problems that have occurred previously.

e) Reference

[1] K. D. P. D. T. dan Transmigrasi, "Indeks Desa Membangun 2015," *Kementrian Desa Pembang. Drh. Tertinggal dan Transm.*, 2016.

[2] 2015 Kementrian Desa. Pembangunan Daerah Tertinggal dan Transmigrasi, "Indeks Desa Membangun 2015," *Kementrian Desa. Pembang. Drh. Tertinggal dan Transm.*, 2015.

[3] P. M. Hasugian, "Pengujian Algoritma Apriori Dengan Aplikasi Weka Dalam," *J. Mantik Penusa*, vol. 1, no. 2, pp. 98–103, 2017.

[4] P. M. Hasugian, "PENGUJIAN ALGORITMA APRIORI DENGAN APLIKASI WEKA DALAM PEMBENTUKAN ASOSIATION RULE," *J. Mantik Penusa*, 2017.

[5] M. Steinbach, G. Karypis, and V. Kumar, "A Comparison of Document Clustering Techniques," *KDD Work. text Min.*, 2000.

[6] F. Murtagh and P. Legendre, "Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?," *J. Classif.*, 2014.

[7] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: An overview," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 2012.

[8] a Moore, "K-means and Hierarchical Clustering," *Stat. Data Min. Tutorials*, 2001.

[9] W. H. E. Day and H. Edelsbrunner, "Efficient algorithms for agglomerative hierarchical clustering methods," *J. Classif.*, 1984.

[10] M. F. Balcan and P. Gupta, "Robust hierarchical clustering," in *COLT 2010 - The 23rd Conference on Learning Theory*, 2010.

[11] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Trans. Neural Networks*, 2000.

[12] G. Y. Wiederschain, "Data mining techniques for the life sciences," *Biochem.*, 2011.

