Vol.6, No.2, December 2023
Available online at: http://jurnal.kdi.or.id/index.php/bt

Evaluation of Distance Measurement Using Complete Linkage Method

Fibia Sentauri Cahyaningrum[1)*], Isna Ayu Safitri Kusuma Dewi[2)], Nola Riwibowo[3)], Taruna Firlian Tama[4)]

[1)2)3)4)]*Institut Teknologi dan Bisnis Ahmad Dahlan Lamongan*
Jl. KH. Ahmad Dahlan 41, Lamongan, Indonesia

[1)]fibiasenta22@gmail.com

[2)]isnayuskd@gmail.com

[3)]n.riwibowo@gmail.com

[4)]tarunafirliantama@gmail.com

Article history:

Received 01 Dec 2023;
Revised 05 Dec 2023;
Accepted 06 Dec 2023;
Available online 28 Dec 2023

Keywords:

Complete Linkage Method
Canberra Metric
Czekanowski Coefficient
Euclidean Distance
Hierarchical Method

Abstract

Cluster analysis is the process of grouping a number of objects based on information obtained from data that explains the relationship between objects with the principle of maximizing similarities between members of one cluster and minimizing similarities between clusters. Cluster analysis is useful for identifying objects (recognition), supporting decision-making systems, and data mining. Cluster analysis consists of hierarchical (Average Linkage, Single Linkage, Complete Linkage, Ward's, and Centroid) and non-hierarchical (K-Means) methods. Each method generally has advantages and disadvantages. Apart from that, there are several distance measures that are commonly used in the grouping process, such as Euclidean, Canberra Metric, Czekanowski Coefficient, and others. In general, researchers will choose one or several cluster analysis methods as a comparison and a certain distance measure to be applied to the data in order to group objects based on certain criteria. In this research, a study and evaluation of Euclidean distance measures, Canberra Metric, and Czekanowski Coefficient were carried out using the Complete Linkage method based on simulated data. The conclusion obtained from evaluating measures of object similarity, namely Euclidean distance, Canberra Metric, and Czekanowski Coefficient by applying the Complete Linkage method, concluded that Euclidean distance is better used as a measure of object similarity in grouping cases compared to Canberra Metric and Czekanowski Coefficient.

I. INTRODUCTION

Cluster analysis is a multivariate technique that aims to group objects based on their similar characteristics. The characteristics of objects in a group have a high level of similarity, while the characteristics of objects in one group and other groups have a low level of similarity. In other words, diversity within a group is minimum, while diversity between groups is maximum [1]–[4].

Cluster analysis methods are divided into two, namely hierarchical methods and non-hierarchical methods. Hierarchical method, is a method that creates a hierarchical decomposition (levels) of a data set or object in a structured manner based on similarities in their properties and the number of desired clusters is not yet known, usually displayed in the form of a dendrogram to make the hierarchy process easier. The examples of hierarchical methods include Single Linkage, Complete Linkage, Average Linkage, Ward’s Method, and Centroid Method. Meanwhile, non-hierarchical methods are used to group objects where the number of clusters to be formed can be determined first as part of the clustering procedure. An example of a non-hierarchical method that is most often used is the K-Means method [5].

In applying cluster analysis, researchers can select relevant variables according to the purpose of grouping. For example, if a researcher wants to know how districts/cities in East Java are grouped based on success in the health sector, the variables that can be used include the Infant Mortality Rate (IMR), the percentage of malnutrition, the percentage of health workers to the total population, and others [6], [7].

In various studies, researchers generally use several cluster analysis methods to apply to data, then choose which method produces the best cluster analysis based on the validity value. In addition, researchers determine

* Corresponding author

ISSN 2622-2728 (online) 2622-271X (print)
doi: 10.32877/bt.v6i2.1045

the distance measure first that will be used in calculating the cluster analysis method without comparing it with other distance measures. Therefore, in this research, a study and evaluation of Euclidean distance measures, Canberra Metric, and Czekanowski Coefficient will be carried out using the Complete Linkage method.

II. RELATED WORKS/LITERATURE REVIEW

Much research has been conducted on the application of various cluster analysis methods to real data. Akmal Fikri, et al have conducted research entitled Comparison of K-Means Clustering and Complete Linkage in Grouping Financial Technology Loan Distribution showing that the K-Means method is able to form an optimal number of clusters better than Complete Linkage based on the DBI value [8]. Yanuwar Reinaldi, et al researched the Comparison of Single Linkage, Complete Linkage, and Average Linkage Methods on Community Welfare in East Java, showing that the Average Linkage method provided the best grouping results based on the Silhouette Index [9]. Susiana Thaib, et al conducted research entitled Application of Hierarchical Cluster Analysis Using Average, Single, and Complete Linkage Methods on COVID-19 Patient Data in Indonesia (Case Study: IHSG Data 2016-2021) showing that the Average Linkage method is the best method for conducting grouping based on Cophenetic correlation values [10].

The research entitled Cluster Analysis for Grouping Provinces in Indonesia based on The Farmer Exchange Rate Sub-sector was carried out by Dwi Amelia, et al using the Single Linkage, Complete Linkage, and Average Linkage methods, showing that the Average Linkage method is the best method based on Cophenetic correlation values [11]. Furthermore, Nurissaidah Ulinnuh, et al conducted research entitled Cluster Analysis in Grouping Provinces in Indonesia Based on Infectious Disease Variables Using the Complete Linkage, Average Linkage, and Ward's Methods showing that the Ward's Method is the best method based on the standard deviation value in the group [12]. Research entitled Comparison of Single Linkage, Complete Linkage, and Average Linkage Methods in Grouping Districts Based on Livestock Type Variables in Sidorjao Regency by Sulthan Fikri Mu'afa, et al shows that the Complete Linkage method is the best method based on the standard deviation ratio Sw to Sb [13].

Based on several studies above, it shows that each data case provides different conclusions regarding the selection of the best cluster method. Therefore, we can conclude that data distribution influences the selection of the best cluster method. Apart from that, the choice of distance measure at the grouping stage also needs to be considered because the results of the distance calculation are thought to influence the results of selecting the best cluster method.

III. METHODS

A. Object Grouping Method

Common methods used in grouping objects are hierarchical methods and non-hierarchical methods. The following is an explanation of the object grouping method.

1) Hierarchical Method

Hierarchical grouping/clustering is used to group objects in a structured manner based on similar properties and the number of groups formed is not yet known. There are two ways to get groups using the hierarchical method, namely merging and separating groups. Merger is obtained by gradually combining objects or groups that are similar so that in the end a new group is obtained. On the other hand, the separation method starts from a large group consisting of all observation objects, then the objects with the highest dissimilarity values are separated, and so on until it is found that diversity within a group is minimum, while diversity between groups is maximum. Examples of hierarchical methods by combining are Single Linkage, Complete Linkage, Average Linkage, and Ward's. Meanwhile, examples of hierarchical methods using separation are Splinter Average Distance and Automatic Interaction Detection (AID) [14].

2) Non-Hierarchical Method

Non-hierarchical grouping/clustering is a method of grouping objects in which the desired number of groups is determined first. The first step of the non-hierarchical cluster method is to select a group as the initial central group and all objects with the closest distance are placed in the cluster that has been formed. The next step is to select a new group and continue placing objects until all objects belong to a group with high similarity.

This research will focus on the hierarchical cluster method, namely the Complete Linkage method.

B. Complete Linkage Method

The Complete Linkage method is a hierarchical cluster method where the distance between clusters is determined from the farthest distance between two objects in different groups. This method can be used well in cases where the objects are from completely different groups.

The first step that must be taken is to calculate the distance between two objects (d_{ij}) as a measure of similarity. The distance measure for calculating the similarity between two objects will be explained in the next sub-chapter. The second step is to find the shortest distance between two objects, then combine the two objects into one new group. For example, the two objects are denoted by group U and group V to get a combined group, namely (UV). Next, to calculate the group distance (UV) with other groups, it is formulated using the following equation (1) [14], [15].

$$d_{(UV)W} = \max(d_{UW}, d_{VW}) \tag{1}$$

where d_{UW} and d_{VW} describe the furthest distance between groups *U* and *W* and *V* and *W*.

C. Measures of Object Similarity

The following is a description of distance measures that are often used to describe similarities between objects.

1) Euclidean

Euclidean is the most commonly chosen distance measure to measure similarity between objects. The Euclidean distance between two objects of dimension p where $\mathbf{x}' = [x_1, x_2, \dots, x_p]$ and $\mathbf{y}' = [y_1, y_2, \dots, y_p]$ is shown in equation (2)

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_p - y_p)^2} \tag{2}$$
$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y})}$$

Euclidean distance is usually calculated from raw data or not from standard data. The advantage of Euclidean distance is that the distance between two objects is not affected by the addition of new objects to be analyzed, which may be outliers. Meanwhile, the disadvantage of Euclidean distance is that the distance can be very large due to differences in scale. For example, if a dimension is measured in cm distance units, and converted in mm (by multiplying the value by 10), the results of the Euclidean distance can be very different, so the results of the cluster analysis may be different.

2) Canberra Metric

Canberra distance is the number of absolute difference values between objects for each variable divided by the number of variable values between objects. This addition continues up to the number of variables used for grouping. This distance is expressed in equation (3)

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=i}^{p} \frac{|x_i - y_i|}{(x_i + y_i)} \tag{3}$$

The Canberra distance measure is only used for variables that have positive values. The advantage of the Canberra distance is that this measure can anticipate if some of the variables used have a long range of values. Meanwhile, the disadvantage of Canberra Metric is that the distance between two objects will be affected when new objects are added which may be outliers.

3) Czekanowski Coefficient

The Czekanowski distance is the difference between one and two times the minimum value of one variable and another variable divided by the sum between the following two variables. This distance can be formulated in equation (4)

$$d(\mathbf{x}, \mathbf{y}) = 1 - \frac{2\sum_{i=1}^{p} \min(x_i, y_i)}{\sum_{i=1}^{p} (x_i + y_i)} \tag{4}$$

The Czekanowski distance measure is also only used for variables that have positive values. The advantages and disadvantages of the Czekanowski Coefficient similarity measure are the same as the Canberra Metric.

D. Best Group Comparison Criteria

One of the best criteria for comparing groups is to use the Agglomeration Coefficient (AC). For example, given $d(i)$ is the distance of the i^{th} object to the first group formed divided by the distance of the i^{th} object to the last formed group. Then the Agglomeration Coefficient is the average of all possible $1 - d(i)$ with $0 < AC < 1$. If AC approaches the value 1, then the grouping formed is getting better, that is, the diversity of objects within the group is minimum, while the diversity between groups is maximum. On the other hand, if AC approaches 0, then the groupings formed are increasingly poor, so that the diversity of objects in the group is large.

IV. RESULTS

A. Example of Calculation of Euclidean Distance, Canberra Metric, and Czekanowski Coefficient

For example, given data consisting of 5 observations with two relevant variables to group these observations. The data is given in Table 1.

TABLE 1
EXAMPLE DATA FOR CALCULATING DISTANCE MEASURESS

Observation	X_1	X_2
A	1	2
B	3	1
C	3	5
D	4	9
E	5	7

Graphically, the observations in Table 1 can be visualized using a scatter plot as in Fig. 1.

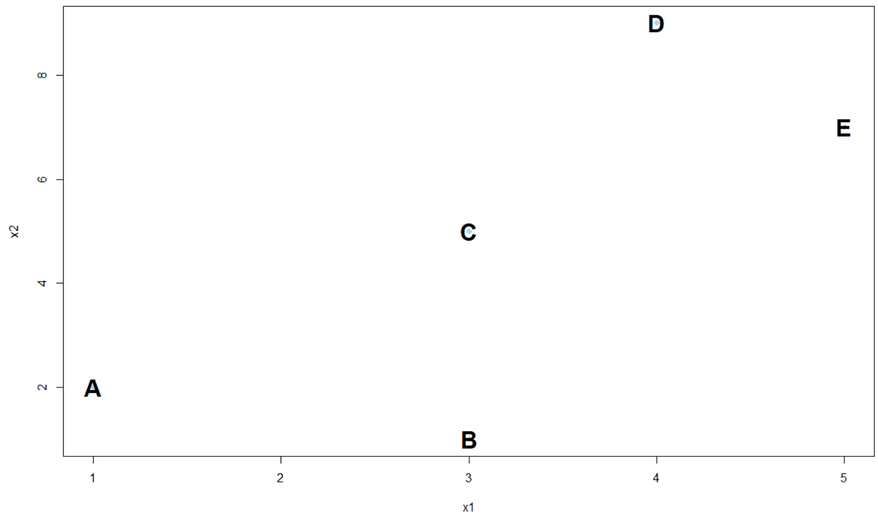

Fig. 1 Scatter Plot of Example Data

Next, we will give examples of how to calculate Euclidean, Canberra Metric, and Czekanowski Coefficient distances between observations.

1) Euclidean Distance

The following is how to calculate Euclidean distance by referring to Equation (2)

$$d_{AB} = \sqrt{(1-3)^2 + (2-1)^2} = \sqrt{5} = 2.236$$
$$d_{AC} = \sqrt{(1-3)^2 + (2-5)^2} = \sqrt{13} = 3.606$$
$$d_{AD} = \sqrt{(1-4)^2 + (2-9)^2} = \sqrt{5} = 7.616$$
$$\vdots$$
$$d_{DE} = \sqrt{(4-5)^2 + (9-7)^2} = \sqrt{5} = 2.236$$

So, the Euclidean distance matrix is obtained as follows

	A	B	C	D	E
A	0	2.236068	3.605551	7.615773	6.403124
B	2.236068	0	4	8.062258	6.324555
C	3.605551	4	0	4.123106	2.828427
D	7.615773	8.062258	4.123106	0	2.236068
E	6.403124	6.324555	2.828427	2.236068	0

2) Canberra Metric

The following is how to calculate the Canberra Metric by referring to Equation (3)

$$d_{AB} = \frac{|1-3|}{(1+3)} + \frac{|2-1|}{(2+1)} = \frac{2}{4} + \frac{1}{3} = 0.833$$

$$d_{AC} = \frac{|1-3|}{(1+3)} + \frac{|2-5|}{(2+5)} = \frac{2}{4} + \frac{3}{7} = 0.929$$

$$d_{AD} = \frac{|1-4|}{(1+4)} + \frac{|2-9|}{(2+9)} = \frac{3}{5} + \frac{7}{11} = 1.236$$

$$\vdots$$

$$d_{DE} = \frac{|4-5|}{(4+5)} + \frac{|9-7|}{(9+7)} = \frac{1}{9} + \frac{2}{16} = 0.236$$

So, the Canberra Metric matrix is obtained as follows

	A	B	C	D	E
A	0	0.833333	0.928571	1.236364	1.222222
B	0.833333	0	0.666667	0.942857	1
C	0.928571	0.666667	0	0.428571	0.416667
D	1.236364	0.942857	0.428571	0	0.236111
E	1.222222	1	0.416667	0.236111	0

3) Czekanowski Coefficient

The following is how to calculate the Czekanowski Coefficient by referring to Equation (4)

$$d_{AB} = 1 - \frac{2\{\min(1,3) + \min(2,1)\}}{(1+3)+(2+1)} = 1 - \frac{2 \times \{1+1\}}{7} = 0.429$$

$$d_{AC} = 1 - \frac{2\{\min(1,3) + \min(2,5)\}}{(1+3)+(2+5)} = 1 - \frac{2 \times \{1+2\}}{11} = 0.455$$

$$d_{AD} = 1 - \frac{2\{\min(1,4) + \min(2,9)\}}{(1+4)+(2+9)} = 1 - \frac{2 \times \{1+2\}}{16} = 0.625$$

$$\vdots$$

$$d_{DE} = 1 - \frac{2\{\min(4,5) + \min(9,7)\}}{(4+5)+(9+7)} = 1 - \frac{2 \times \{4+7\}}{25} = 0.12$$

So, the Czekanowski Coefficient distance matrix is obtained as follows

	A	B	C	D	E
A	0	0.428571	0.454546	0.625	0.6
B	0.428571	0	0.333333	0.529412	0.5
C	0.454546	0.333333	0	0.238095	0.2
D	0.625	0.529412	0.238095	0	0.12
E	0.6	0.5	0.2	0.12	0

B. Example of Grouping Object with the Complete Linkage Method

After calculating the distance between objects using various similarity measures, namely Euclidean, Canberra Metric, and Czekanowski Coefficient, the next step is to group objects based on their similarity using the Complete Linkage method. The following will give an example of how to group objects using the Complete Linkage method with the Euclidean distance measure.

1) The first stage is to determine the smallest d_{ij} from the Euclidean distance matrix

	A	B	C	D	E
A	0	2.236068	3.605551	7.615773	6.403124
B	2.236068	0	4	8.062258	6.324555
C	3.605551	4	0	4.123106	2.828427
D	7.615773	8.062258	4.123106	0	2.236068
E	6.403124	6.324555	2.828427	**2.236068**	0

In the first stage, objects D and E are combined into one group because the two objects have the highest similarity which is indicated by the minimum distance between the two objects.

2) The second stage is to determine the distance of objects A, B, and C to the group (DE) based on equation (1), which is as follows

$$d_{(DE)A} = \max\{d_{DA}, d_{EA}\} = \max\{7.615773, 6.403124\} = 7.615773$$

$$d_{(DE)B} = \max\{d_{DB}, d_{EB}\} = \max\{8.062258, 6.324555\} = 8.062258$$

$$d_{(DE)C} = \max\{d_{DC}, d_{EC}\} = \max\{4.123106, 2.828427\} = 4.123106$$

so that a new distance matrix is obtained as follows

	(DE)	A	B	C
(DE)	0	7.615773	8.062258	4.1231061
A	7.615773	0	2.236068	3.605551
B	8.062258	**2.236068**	0	4
C	4.123106	3.605551	4	0

3) In the third stage, a new group is formed, namely the combination of objects A and B because the distance between the two objects is minimum so that group (AB) is obtained. Next, the distance between (AB), (DE), and C is calculated based on Equation (1), as follows

$$d_{(AB)(DE)} = \max\{d_{A(DE)}, d_{B(DE)}\} = \max\{7.615773, 8.062258\} = 8.062258$$

$$d_{(AB)C} = \max\{d_{AC}, d_{BC}\} = \max\{3.605551, 4\} = 4$$

so that a new distance matrix is obtained as follows

	(DE)	(AB)	C
(DE)	0	8.062258	4.123106
(AB)	8.062258	0	4
C	4.123106	**4**	0

After calculating the new distance matrix, it is found that the minimum distance occurs between object C and group (AB) so that object C joins group (AB) to become group (CAB).

4) The fourth stage is to create a new distance matrix between groups (CAB) and (DE) based on Equation (1).

$$d_{(CAB)(DE)} = \max\{d_{C(DE)}, d_{(AB)(DE)}\} = \max\{4.123106, 8.062258\} = 8.062258$$

so that a new distance matrix is obtained as follows

	(DE)	(CAB)
(DE)	0	8.062258
(CAB)	8.062258	0

The new matrix obtained in stage four shows that the grouping of objects in Table 1 forms 2 large groups where group 1 consists of objects D and E, and group 2 consists of objects A, B and C. The results of this grouping can be shown visually using the dendrogram in Fig. 2 below

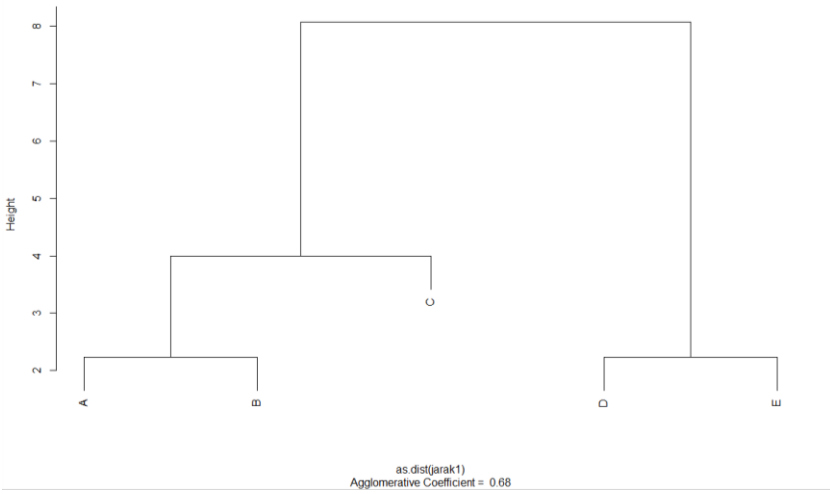

Fig. 1 Complete Linkage Dendrogram Using Euclidean Distance

The stages of grouping objects with the Canberra Metric and Czenowski Coefficient similarity measures using the Complete Linkage method are the same. With the help of computing, grouping results are obtained as in Fig. 2 and Fig. 3.

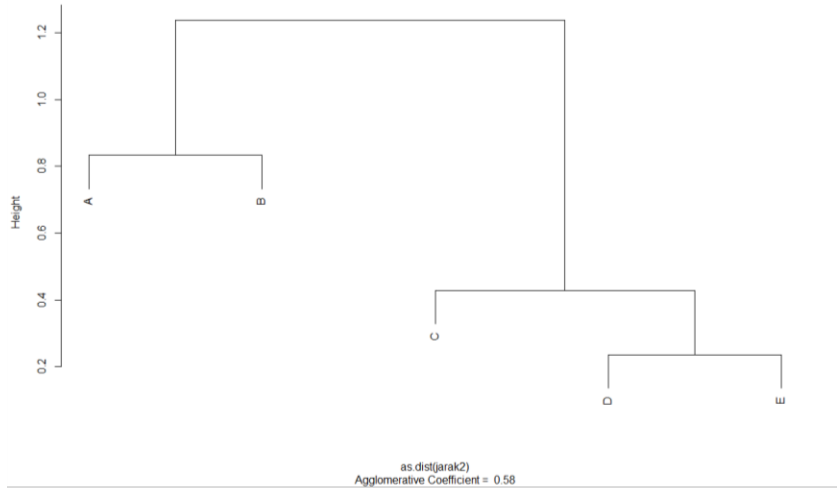

Fig. 2 Complete Linkage Dendrogram Using Canberra Metric Distance

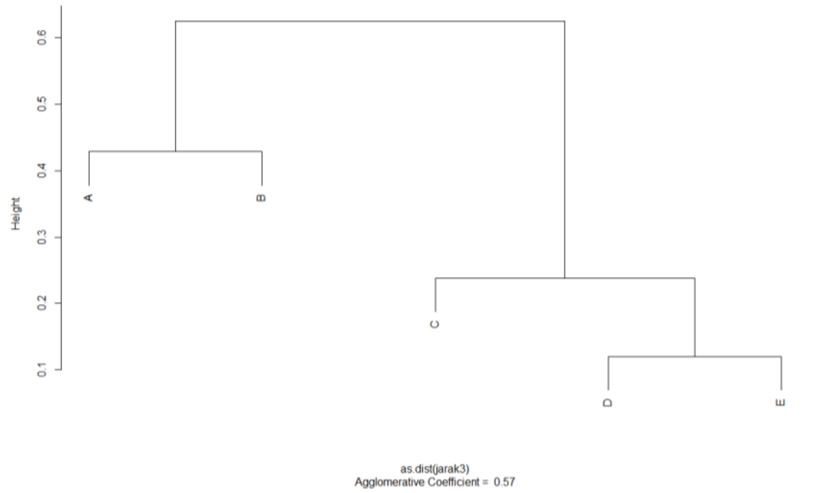

Fig3. Complete Linkage Dendrogram Using Czekanowski Coefficient Distance

The results of grouping objects using the Complete Linkage method, which measured their similarity using the Canberra Metric distance and the Czenowski Coefficient, are the same. However, the results of these two groupings are different from object groupings when using Euclidean distance as a measure of object similarity. To find out the best similarity measure in grouping objects in Table 1, an evaluation was carried out using the Agglomeration Coefficient (AC) measure. With the help of computing, the AC for Euclidean distance, Canberra Metric, and Czenowski Coefficient are 0.6788921, 0.5846755, and 0.5727238 respectively. The conclusion obtained from this calculation is that Euclidean distance is more used to group objects (with the Complete Linkage method) than Canberra Metric and Czenowski Coefficient.

C. Evaluation of Object Similarity Measures by Applying the Complete Linkage Method

In the previous sub-chapter, we explained the stages of grouping objects using similarity measures, namely Euclidean distance, Canberra Metric, and Czekanowsi Coefficient along with evaluating the grouping results. Based on the example in the previous sub-chapter, it was found that Euclidean distance is the best measure of similarity as indicated by the maximum Agglomeration Coefficient value.

To validate whether the conclusions from the evaluation using the example data above are correct, in this sub-chapter an evaluation will be carried out on the Euclidean distance, Canberra Metric, and Czekanowsi Coefficient. This evaluation was carried out by generating data with the number of observations being 10 which were repeated 100 times with the number of variables being 2, then the 100 datasets calculated the similarity measure using Euclidean distance, Canberra Metric, and Czekanowsi Coefficient and continued with grouping objects using the Complete Linkage method. After the grouping is carried out, it is continued by calculating the Agglomeration Coefficient for each grouping, so that 100 Agglomeration Coefficients are obtained for each distance. The way to determine the best distance is by ranking the Agglomeration Coefficient. The smallest ranking is given to the smallest Agglomeration Coefficient, then the ranking results are added up. The best measure of object/distance similarity is indicated by the largest number of rankings. Apart from that, the evaluation is also based on the average Agglomeration Coefficient for each object similarity/distance measure.

Table 2 below shows the Agglomeration Coefficient for each iteration for various object similarity/distance measures.

TABLE 2
EVALUATION RESULT OF OBJECT SIMILARITY MEASURES

Iteration	*Euclidean*	*Canberra*	*Czekanowski*	*Euclidean* Rank	*Canberra* Rank	*Czekanowski* Rank
1	0.6514446	0.6826444	0.6826444	1	2.5	2.5
2	0.7915499	0.7495870	0.749587	3	1.5	1.5
3	0.6959237	0.7508384	0.7508384	1	2.5	2.5
4	0.6966339	0.6916998	0.6916998	3	1.5	1.5
5	0.8088821	0.7249177	0.7249177	3	1.5	1.5
⋮	⋮	⋮	⋮	⋮	⋮	⋮
98	0.7146157	0.6828146	0.6828146	3	1.5	1.5
99	0.8321283	0.7395675	0.7395675	3	1.5	1.5
100	0.7771369	0.7013668	0.7013668	3	1.5	1.5

V. DISCUSSION

The sum of Euclidean distance rankings, Canberra Metric, and Czekanowsi Coefficient are 268, 166, and 166 respectively. Meanwhile, the average Agglomeration Coefficient of grouping objects using these distances is 0.7647577, 0.7306839, and 0.7306839, respectively. Based on the simulation, it can be concluded that Euclidean distance is better used as a measure of object similarity compared to Canberra Metric and Czekanowski Coefficient distances. The results of grouping objects using the Canberra Metric and Czekanowski Coefficient distances tend to be the same and the results of grouping distances using these two measures tend to be different from the results of grouping using Euclidean distance. This tendency means that the grouping results between the three object similarity measures are the same.

Therefore, the Euclidean distance is very popular for use as a measure of object similarity in group/cluster analysis. Apart from that, the advantage of the Euclidean distance compared to the Canberra Metric distance and the Czekanowski Coefficient is that the Euclidean distance can be used for variables with negative values, while the Canberra Metric distance and the Czekanowski Coefficient cannot be used for variables with negative values because the Canberra Metric formula contains the absolute difference in variable values between observations. and the Czekanowski Coefficient formula looks for the minimum value of a variable between objects, so that if these two distances are applied to variables with negative values, inconsistent results will be obtained. Apart from that, if a new object is added which may be an outlier, it does not affect the Euclidean distance, but it does affect the Canberra Metric distance and Czekanowski Coefficient.

VI. CONCLUSIONS

The conclusion contains a summary of what is learned from the results obtained, what needs to be improved in further study. Other common features of the conclusions are the benefits and applications of the research, limitation, and the recommendations based on the results obtained.

The conclusion obtained from evaluating measures of object similarity, namely Euclidean distance, Canberra Metric, and Czekanowski Coefficient by applying the Complete Linkage method, concluded that Euclidean distance is better used as a measure of object similarity in grouping cases compared to Canberra Metric and Czekanowski Coefficient.

Suggestions that can be given for further evaluation are as follows: (1) Evaluation should be carried out by combining grouping methods, such as Single Linkage, Average Linkage, etc; (2) The criteria for comparing groups should not only use the Agglomeration Coefficient; (3) It is best to carry out simulations for various numbers of variables and the number of objects used in one iteration; (4) It is recommended that simulated generation data be created in three scenarios, namely visually objects have formed very clear groups, visually objects have formed unclear groups, and visually objects have not formed unclear groups. With a design according to the suggestions, a valid conclusion can be obtained.

REFERENCES

[1] "Practical Multivariate Analysis - Abdelmonem Afifi, Susanne May, Robin Donatello, Virginia A. Clark - Google Buku." https://books.google.co.id/books?hl=id&lr=&id=D5K3DwAAQBAJ&oi=fnd&pg=PP1&dq=multivariate+analysis&ots=vU-LkLdOhc&sig=oIjlZAGGIVWQU9Pc6b6dx5kGm44&redir_esc=y#v=onepage&q=multivariate analysis&f=false (accessed Dec. 05, 2023).

[2] "Approaching Multivariate Analysis, 2nd Edition: A Practical Introduction - Pat Dugard, John Todman, Harry Staines - Google Buku." https://books.google.co.id/books?hl=id&lr=&id=5hh1EAAAQBAJ&oi=fnd&pg=PP1&dq=multivariate+analysis&ots=IsLcpTLwkJ&sig=tqhp_THm7ZOOp6O4NanF53HlyfE&redir_esc=y#v=onepage&q=multivariate analysis&f=false (accessed Dec. 05, 2023).

[3] "Model-Based Clustering and Classification for Data Science: With ... - Charles Bouveyron, Gilles Celeux, T. Brendan Murphy, Adrian E. Raftery - Google Buku." https://books.google.co.id/books?hl=id&lr=&id=ldGoDwAAQBAJ&oi=fnd&pg=PR15&dq=multivariate+analysis+of+cluster+analysis+complete+linkage&ots=6Mh5-zLn4n&sig=jrMIWAMC5lT1jXA8DnUzdDtpFJU&redir_esc=y#v=onepage&q=multivariate analysis of cluster analysis complete linkage&f=false (accessed Dec. 05, 2023).

[4] A. Kassambara, "Multivariate Analysis I Practical Guide To Cluster Analysis in R Unsupervised Machine Learning", Accessed: Nov. 20, 2023. [Online]. Available: http://www.sthda.com

[5] B. S. Everitt, S. Landau, and M. Leese, "Cluster Analysis Arnold," *A Memb. Hodder Headl. Group, London*, p. 237, 2001, Accessed: Nov. 20, 2023. [Online]. Available: https://books.google.com/books/about/Cluster_Analysis.html?hl=id&id=htZzDGlCnQYC

[6] R. Scitovski, K. Sabo, F. Martínez-Álvarez, and Š. Ungar, "Cluster Analysis and Applications," *Clust. Anal. Appl.*, pp. 1–271, Jan. 2021, doi: 10.1007/978-3-030-74552-3/COVER.

[7] A. Jaeger and D. Banks, "Cluster analysis: A modern statistical review," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 15, no. 3, p. e1597, May 2023, doi: 10.1002/WICS.1597.

[8] A. Fikri, B. F. Hutabarat, and U. Khaira, "Komparasi K-Means Clustering Dan Complete Linkage Dalam Pengelompokan Penyaluran Pinjaman Financial Technology," *J. Ilm. Media Sisfo*, vol. 17, no. 2, pp. 228–239, Oct. 2023, doi: 10.33998/MEDIASISFO.2023.17.2.1373.

[9] Y. Reinaldi, N. Ulinnuha, T. Hartono, and M. Hafiyusholeh, "Comparison of Single Linkage, Complete Linkage, and Average Linkage Methods on Community Welfare Analysis in Cities and Regencies in East Java," *J. Mat. Stat. dan Komputasi*, vol. 18, no. 1, pp. 130–140, Sep. 2021, doi: 10.20956/J.V18I1.14228.

[10] S. Thaib and R. D. Bekti, "PENERAPAN ANALISIS KLUSTER HIERARKI MENGGUNAKAN METODE AVERAGE, SINGLE, DAN COMPLETE LINKAGE PADA DATA PASIEN COVID-19 DI INDONESIA (Studi Kasus : Data IHSG Tahun 2016 – 2021)," *J. Stat. Ind. dan Komputasi*, vol. 7, no. 2, pp. 23–30, 2022, Accessed: Nov. 20, 2023. [Online]. Available: https://ejournal.akprind.ac.id/index.php/STATISTIKA/article/view/4389

[11] D. Amelia, G. Kholijah, F. Sains dan Teknologi, and U. Jambi, "Analisis Cluster Pengelompokan Provinsi di Indonesia Berdasarkan Sub Sektor Nilai Tukar Petani," *J. Demogr. Ethnogr. Soc. Transform.*, vol. 3, no. 1, pp. 1–12, Jun. 2023, doi: 10.30631/DEMOS.V3I1.1812.

[12] N. Ulinnuha and R. Veriani, "Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage dan Ward," *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 5, no. 1, pp. 102–108, Sep. 2020, doi: 10.30743/INFOTEKJAR.V5I1.2464.

[13] S. F. Mu'afa and N. Ulinnuha, "Perbandingan Metode Single Linkage, Complete Linkage Dan Average Linkage dalam Pengelompokan Kecamatan Berdasarkan Variabel Jenis Ternak Kabupaten Sidoarjo," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 4, no. 2, 2019, doi: 10.25139/inform.v4i2.1696.

[14] A. L. Scutariu, Ștefăniță Șușu, C. E. Huidumac-Petrescu, and R. M. Gogonea, "A Cluster Analysis Concerning the Behavior of Enterprises with E-Commerce Activity in the Context of the COVID-19 Pandemic," *J. Theor. Appl. Electron. Commer. Res. 2022, Vol. 17, Pages 47-68*, vol. 17, no. 1, pp. 47–68, Dec. 2021, doi: 10.3390/JTAER17010003.

[15] P. Giordani, M. B. Ferraro, and F. Martella, "An Introduction to Clustering with R," vol. 1, 2020, doi: 10.1007/978-981-13-0553-5.