Journal of Theoretical and Applied Information Technology
15th June 2021. Vol.99. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

THE OPTIMAL CUSTER BASED ON COMBINATORIAL OPTIMIZATION APPROACH FOR DATA DETERMINATION ALGORITHM IN CLUSTER

[1]DENY JOLLYTA, [2*]SYAHRIL EFENDI, [3]MUHAMMAD ZARLIS, [4]HERMAN MAWENGKANG

[1]Graduate Program of Computer Science, Department of Computer Science, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia

[2,3]Department of Computer Science, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia

[4]Department of Mathematics, Faculty of Mathematics, Universitas Sumatera Utara, Medan, Indonesia

E-mail: [1]deny.jollyta@student.usu.ac.id, [2*]syahril1@usu.ac.id, [3]m.zarlis@usu.ac.id, [4]mawengkang@usu.ac.id

ABSTRACT

Clustering still leaves problems in selecting optimal clusters in order to obtain a right and correct classification analysis. Right in the sense of the number of clusters, while correct in terms of the information generated by a group of cluster members that is optimally grouped. Determining the optimal number of clusters is a difficult problem in non-polynomials. A number of existing approaches generally still rely on the number of K tests tested. This study aims to produce a new approach that can determine and place data in clusters optimally in a combinatorial form. This can be done by considering that the problem of selecting cluster placement has a combinatorial optimization structure pattern. However, the resulting combinatorial optimization model is quadratic. Therefore, in order to make the combinatorial clustering problem easier to solve, linearization of the cluster data was carried out so that a combinatorial optimization approach was produced with the algorithm. Several illustrations have been put forward to demonstrate the validity of the method. The combinatorial optimization approach as proposed in this research produces novelty on cluster data analysis techniques.

Keywords: *Clustering, Information, Combinatorial Optimization, Linearization, Cluster Data*

1. INTRODUCTION

Clustering and determining the number of clusters are included in an unlimited form of operation. Clusters can be formed from a little or a lot of data with few or many criteria so that it is necessary to reorganize the data before using it through the stages arranged in the Knowledge Discovery in Database (KDD). This is possible because generally the information that is expected to come from data with certain criteria or just desired [1].

Determining and placing the optimal amount of data in clusters is a difficult problem in non-polynomials. This difficulty level is increasing with the problem of converting the clustering problem into a combinatorial problem, so we need a new approach that can determine the optimal number of N objects in the cluster in a combinatorial form.

In many applications, the similarity that is processed by the evaluation technique is carried out on clusters that have been formed from a number of k tests so that they ignore previously formed cluster members. As a result, even though the optimal cluster has been selected, it is not certain that the data which is a cluster member is also optimal. In research in obtaining this new optimization approach, cluster evaluation techniques are needed to compare with the final results of the approaches obtained.

Determination of clusters and data in clusters using a combinatorial optimization

approach has never been done. A number of existing studies use non-combinatorial evaluation techniques, such as research [2] which uses the Elbow method to determine the optimal cluster through the eK-NNclus algorithm clustering. Dynamic Cluster Algorithm is a solution in optimizing the radius to the center in a group of data matrices [3] and dynamic programs which are the optimal cluster formation solutions for sequential data [4]. In addition, studies [5],[6] solve cluster optimization problems using the Silhouette Index.

A number of studies above showed that cluster optimization techniques in solving clustering problems emphasize the use of a number of k tests that are determined at the beginning and the closest distance in placing cluster members (N objects) in the cluster. Research [7], described that the combinatorial approach allows the placement of N objects in a number of clusters and optimization of the number of objects in a cluster through a combinatorial approach design and the number of clusters itself also depends on the application used. Research conducted by researchers at this time is to produce a new approach that can determine and place data or objects in the cluster optimally in the form of a combinatorial optimization model. Combinatorial solution is shown in the number of N objects that can be generated and occupy the cluster optimally, as well as forming a placement algorithm according to the approach obtained. It is hoped that the results of this study will be an alternative for cluster evaluation techniques in producing the best information from an optimal clustering process.

2. LITERATURE REVIEW

2.1 Clustering

Clusters can be presented in various forms [1]. It really depends on the variety of data being grouped and the algorithm used [8]. Clusters are used in many areas of research such as data mining, statistical data analysis, machine learning, pattern recognition, image analysis and information retrieval. The clustering problem cannot be solved by one particular algorithm but requires a variety of algorithms that differ significantly in terms of what makes up clusters and how to find them efficiently. Generally, clusters include groups with small distances between cluster members, dense areas of data space, certain intervals or statistical distributions. Therefore, grouping can be formulated as multi-objective optimization problems. The appropriate clustering algorithm and parameter setting depend on the individual data set and the intended use of the results. Such cluster analysis is an iterative process of knowledge discovery or interactive multi-purpose optimization that involves trial and error. Often it is necessary to modify preprocessing data and model parameters until the results reach the desired objectives [9].

2.2 Cluster Evaluation Techniques and Data Placement

The difficulty in determining the knowledge / information of clusters, encourages the creation of various algorithms, techniques or methods that can determine the optimal number of clusters. Various evaluation techniques have been developed with different results. A description of some of them is shown in the following table.

Table 1: Cluster Evaluation Technique

Cluster Evaluation Techniques	Description	Research
Elbow Method	- Known as the Elbow method - Number of clusters based on distortion, the average distance per dimension between each nearest cluster center. - The calculation is done using the Sum of Square Error (SSE) equation. - Tested on a number of k - If the percentage value of a cluster with the next cluster has the largest decrease, then the cluster is selected as the optimal cluster	[10]
Information Theoritic Approach	Determine the number of clusters based on distortion, which is a number that measures the average per-dimensional distance between each observation result, the average distance and the center of the closest cluster	[11], [12], [13], [14].
Dynamic Cluster Algorithm	- This algorithm allows nodes to gradually build simple views and run a grouping algorithm to build a	[15], [16], [3], [17]

	grouping model - Fulfillment of the best cluster validity boundary conditions indicates that a cluster is optimal	
Silhouette Index	- The way it works is to interpret and produce consistent validation within a data set - Clusters with an index close to 1 are the best or optimal	[18]
Davies Bouldin Index (DBI)	- This is an internal evaluation scheme, where validation of how well the grouping has been done is made using the quantity and features attached to the data set - DBI is based on the ratio between the distance "within-cluster" and "between-cluster" - The optimal cluster is obtained from the smallest DBI value from a number of k tests	[19]

In many applications, the similarity that is processed by the evaluation technique is carried out on clusters that have been formed from a number of k tests so that they ignore previously formed cluster members. As a result, even though the optimal cluster has been selected, it is not certain that the data that is a cluster member is also optimal. In order to discover this new optimization approach, existing cluster evaluation techniques are needed to compare against the final results of the approaches obtained.

2.3 Combinatorial Optimization Approach Combinatorial

Optimization is used side by side with a method or algorithm [20]. This approach can be applied to various fields to solve various problems, including data mining. In the study [21], combinatorial optimization was applied to the selection of the smallest network size without reducing the capacity size required by Synchronous Optical Network (SONET) subscribers. The results were obtained through the analysis of Capacitated Vehicle Routing Problem (CVRP).

Combinatorial optimization is also useful for project selection using Genetic Algorithms to simplify the project selection process [22] and reduce the number of variables and the cost of survey sample space [23]. In addition, combinatorial optimization can optimize the grouping of census data based on 6 demographic attributes [24] and find new algorithms in the Quadratic Assignment Problem [25]. Meanwhile, the current research uses combinatorial optimization on linearized non-linear problems so that a new approach is needed to produce optimal grouping.

3. RESEARCH METHODOLOGY

To obtain the formulation of a combinatorial optimization approach, a systematic research stage was compiled as shown in Figure 1.

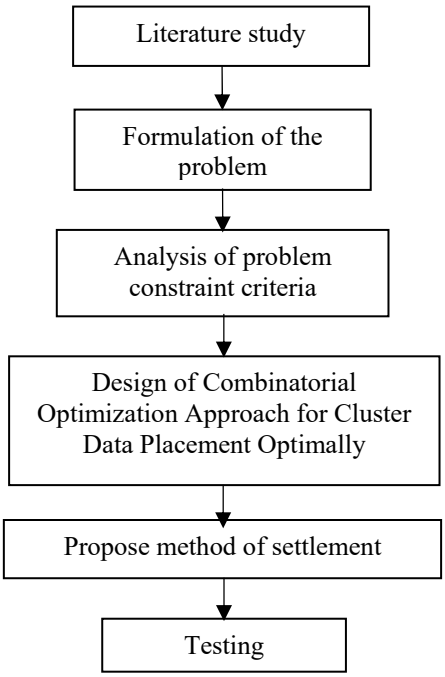

Figure 1: Research Stages

The research begins with enrichment of knowledge about cluster evaluation techniques in determining the optimal cluster. The limitations of each existing technique are formulated in the form of a problem. The formulation in question is to change the problem to a combinatorial form, such

as formulating objects, clusters, distances and determining constraint functions that are calculated to obtain similarity through distance calculations. Furthermore, the design of an approach model is carried out for optimizing the number of clusters and placing the data in the cluster.

After the model is formed, testing is carried out using object simulations with the specified number of objects and clusters. In this study, simulations were carried out for the same number of objects and clusters, as well as for the number of objects and different clusters.

The last stage as shown in Figure 1, the model is tested on a number of data with each constraint. The data are arranged in the form of an NxM matrix and for the constraints X = 1 and X≥1. After the matrix is arranged correctly, the model is tested and executed using the Linear Interactive and Discreate Optimizer (LINDO) application. The resulting output is in the form of integer decisions 0 and 1, where 0 means that the data is not selected as optimal to be placed in the cluster, while 1 means that the data is selected to be placed in the cluster and the placement is the most optimal.

4. RESULT AND DISCUSSION

4.1 Problem Formulation

Problem in determining the optimal number of clusters is the direct emergence of the number of clusters. In this condition, the problem criterion which is a constraint function is calculated to obtain similarity. Similarities are obtained from calculating the distance between objects. The distance in question is from minimizing the groups in a number of objects or data and minimizing the maximum distance in the group. Data grouping is simulated in Figure 2.

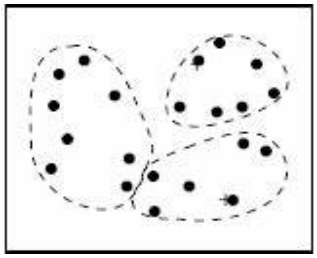

Figure 2: Cluster Illustration

In Figure 2 each object has N constraints. The position of the object becomes the point of calculating the distance between objects. Every object with the same similarity or approaching will automatically be approached to form a group. Formulation of the problem can be given as follows:

- ✓ *A* set of *N* object, where $O=\{O_1, O_2,...O_n\}$
- ✓ *A* set of *M* of pre-assigned cluster, where $S=\{S_1, S_2,..., SM\}$

The distance function *d*, is calculated to determine the distance of each object based on its similarity, where:

- ✓ $d_{ij}>=0$, which means that the object is one distance from another object;
- ✓ $d_{ij}=0$, which means the objects have the same distance;
- ✓ $d_{ij}=d_{ji}$, which means the distance O_1 to O_2 is the same as the distance O_2 to O_1, for $i,j = \{1,...,N\}$.

Thus, to obtain the decision variable, it is shown from the number of *NxM* with the decision $X_{ij} \in \{0,1\}$ so that in the form of a notation it is written:

$$X_{ij} = \begin{cases} 1 \text{ Jika objek } i \text{ berada pada cluster } j \\ 0 \text{ Jika tidak} \end{cases}$$

4.2 Combinatorial Optimization Model

Based on the explanation above, the optimal number of clusters in a combinatorial problem is determined by the cluster distance being linearized. Data clustering (DC) can be formulated as non-linear 0-1 problems as follows:

$$(DC) = \min \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij} \sum_{k=1}^{M} X_{ik}, X_{jk} \quad (1)$$

s, t

$$\sum_{k=1}^{M} X_{ik} = 1, i = 1,..., N \quad (2)$$

$$\sum_{i=1}^{M} X_{ik} \geq 1, k = 1,..., M \quad (3)$$

Where: $X_{ik} \in \{0,1\}, i = 1,..., N, k = 1,..., M$

That is, cluster data is the scope of non-linear problems, where formula (1) can Minimizing the distance between objects in the same cluster, formula (2) can guarantee that each object only occupies a cluster, formula (3) can guarantee that each cluster within the highest test boundary has at least one object.

Improvements to non-linear formulations require a linearization process, in which the linearization function is able to guarantee that

objects with high similarities or having the lowest distance will be in the same cluster, where $\forall_{i,j} = 1,..., N$, then y_{ij}= 1 for O_i, $O_j \in O$ in the same cluster. The full explanation of cluster data linearization (LDC) is as follows.

$$(LDC) = \min \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}, y_{ij} \quad (4)$$

Where can minimize the distance between objects in the same cluster. For the formulation of constraints as shown in equations (5) and (6), the guarantees provided are the same as equations (2) and (3), namely:

s, t

$$\sum_{k=1}^{M} X_{ik} = 1, i = 1,..., N \quad (5)$$

$$\sum_{i=1}^{M} X_{ik} \geq 1, k = 1,..., M \quad (6)$$

$$X_{ik} \in \{0,1\}, i = 1,..., N, k = 1,..., M \quad (7)$$

$$Y_{ij} \geq X_{ik} + Y_{jk} - 1, i = 1,..., N, \; j = i+1,..., N, k \quad (8)$$

$$Y_{ij} \geq 0, i = 1,..., N. j = i+1,..., N \quad (9)$$

Equation (7) is a decision on the position of the cluster and data on the cluster through variable X_{ij}. The variable X, which has a value of 1, indicates the optimal cluster, while the one with 0 is the opposite. Equations (8) and (9) ensures that the Y_{ij}=1 if X_{ik}=X_{jk}=1, where O_i, $O_j \in O$, contained in the same cluster. This happens because LDC has $\frac{N^2}{2}$ movable variables and has $\frac{N(N-1)(M+1)}{2}$ constraints compared to DC, but it makes linearization easier.

4.3 Implementation and Testing

Based on the model design that has been obtained, the model is tested on the same number of objects and clusters (N = M) and different (N> M).

4.3.1 Number of Objects = Number of Clusters (N = M)

In the design to produce the desired combinatorial optimization equation, suppose there are 10 nodes symbolized by N and 10 clusters symbolized by M. The coefficients of all Y represent the distance from the object. By using equation (4) which is linearized from equation (1), the minimum function is arranged in LINDO as follows:

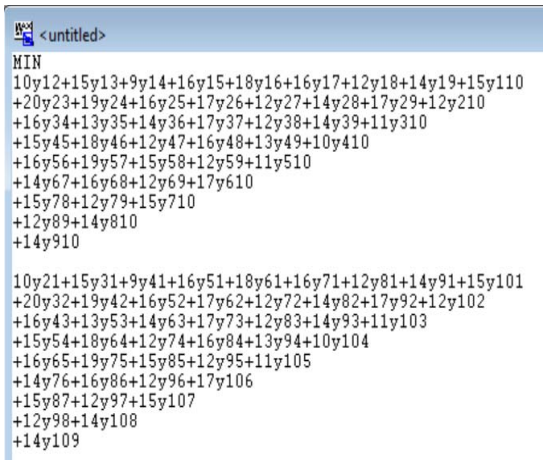

Figure 3: Minimum Function for N=10 M=10

Constraint function is described by the number of x formed from equations (2) and (3) as follows:

Figure 4: Constraint for N=10 M=10

Based on Figure 4, the description of this minimum function becomes the goal to be achieved. The obstacle in the optimization problem is the direct emergence of the number of clusters. The constraint is represented by a number of X.

Journal of Theoretical and Applied Information Technology
15th June 2021. Vol.99. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Then the X and Y values are written as Figure 5, 6 and 7..

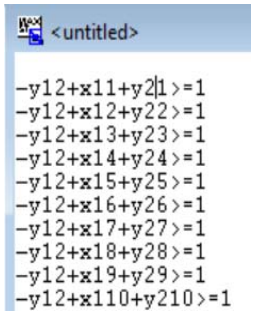

Figure 5: Y Values for N=10 M=10

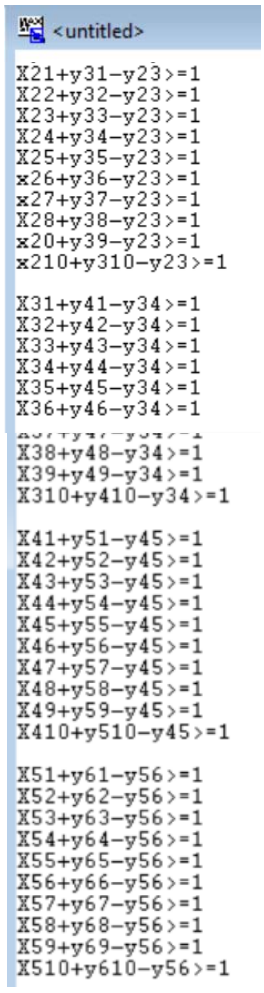

Figure 6: X Values for N=10 M=10

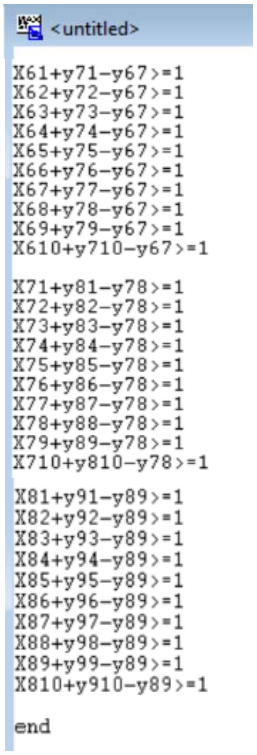

Figure 7: X Values for N=10 M=10 Continued

Where:

$X_{ik} \in \{0,1\} \quad \forall i \in N, \forall k \in M$

$Y_{ij} \geq 0, \quad \forall i \in N, \forall j \in M$

The minimum functions and constraints are tested using the LINDO application Execution result by LINDO showed the value of *X*, as shown in Table 2.

Table 2: Optimal Results for Variable X N=10 M=10

Variable	Value
X11	0,000000
X12	0,000000
X13	1,000000
X14	0,000000
X15	0,000000
X16	0,000000
X17	0,000000
X18	0,000000
X110	0,000000
X21	0,000000
X22	0,000000
X23	0,000000
X24	1,000000
X25	0,000000
X26	0,000000
X27	0,000000
X28	0,000000
X29	0,000000

X210	0,000000
X31	0,000000
X32	0,000000
X33	0,000000
X34	0,000000
X35	1,000000
X36	0,000000
X37	0,000000
X38	0,000000
X39	0,000000
X310	0,000000
X41	0,000000
X42	0,000000
X43	0,000000
X44	0,000000
X45	0,000000
X46	1,000000
X47	0,000000
X48	0,000000
X49	0,000000
X410	0,000000
X51	0,000000
X52	0,000000
X53	0,000000
X54	0,000000
X55	0,000000
X56	0,000000
X57	1,000000
X58	0,000000
X59	0,000000
X510	0,000000
X61	0,000000
X62	0,000000
X63	0,000000
X64	0,000000
X65	0,000000
X66	0,000000
X67	0,000000
X68	1,000000
X610	0,000000
X71	0,000000
X72	0,000000
X73	0,000000
X74	0,000000
X75	0,000000
X76	0,000000
X77	0,000000
X78	0,000000
X79	1,000000
X710	0,000000
X81	0,000000
X82	1,000000
X83	0,000000
X84	0,000000
X85	0,000000
X86	0,000000
X87	0,000000
X88	0,000000
X89	0,000000
X810	0,000000
X91	0,000000
X92	0,000000
X93	0,000000
X94	0,000000
X95	0,000000
X96	0,000000
X97	0,000000
X98	0,000000
X99	0,000000
X910	1,000000
X101	1,000000
X102	0,000000
X103	0,000000
X104	0,000000
X105	0,000000
X106	0,000000
X107	0,000000
X108	0,000000
X109	0,000000
X1010	0,000000
X19	1,000000
X69	1,000000
X21	1,000000

4.3.2 Number of Objects> Number of Clusters (N> M)

The next test is simulated on the number of objects and clusters that are not the same, which are N = 7 and M = 5, N = 10 and M = 6.

For N = 7 and M = 5. minimum function is as follows:

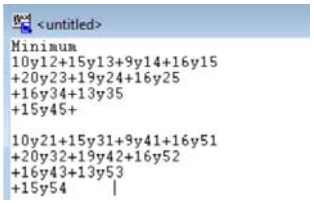

Figure 8: Minimum Function for N=7 M=5

The constraint for *N*=7 and *M*=5 is described by the number of x formed as follows:

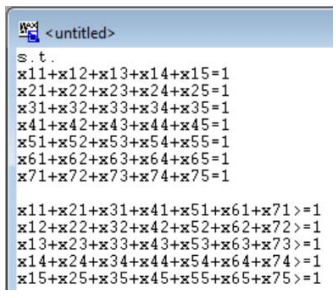

Figure 9: Constraint for N=7 M=5

Next is the writing of the X and Y values. In the application of the model to N = M, equations (2) and (3) support to achieve the desired results. However, in implementation, the number of N is not always the same as M. Some objects can be members of a cluster or N>M. For this reason, the model execution in LINDO can be presented using int function. The X and Y values are shown in Figure 10.

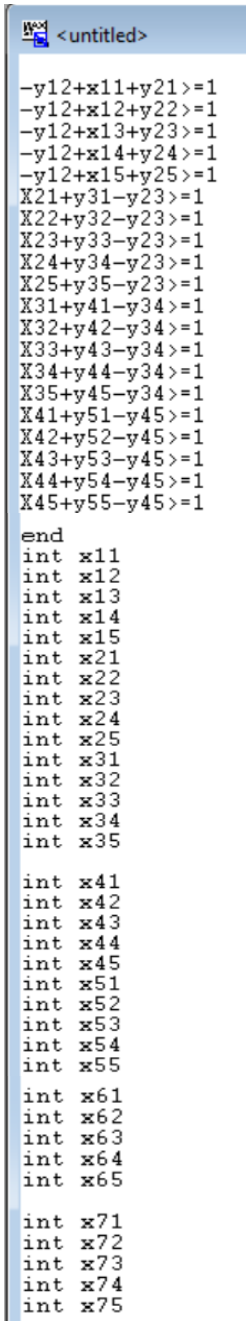

Figure 10: X and Y Values for N=7 M=5

Execution result by LINDO showed the value of *X*, presented in Table 3.

Table 3: Results of Optimal Variable X N=7 M=5

Variable	Value
X11	0,000000
X12	0,000000
X13	1,000000
X14	0,000000
X15	0,000000
X21	0,000000
X22	0,000000
X23	0,000000
X24	1,000000
X25	0,000000
X31	0,000000
X32	0,000000
X33	0,000000
X34	0,000000
X35	1,000000
X41	1,000000
X42	0,000000
X43	0,000000
X44	0,000000
X45	0,000000
X51	0,000000
X52	1,000000
X53	0,000000
X54	0,000000
X55	0,000000
X61	0,000000
X62	0,000000
X63	1,000000
X64	0,000000
X65	0,000000
X71	0,000000
X72	0,000000
X73	1,000000
X74	0,000000
X75	0,000000

For N = 10 and M = 6. The resulting minimum function and constraint of this example are shown in Figure 11 and 12:

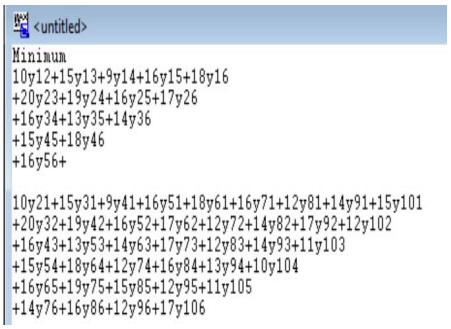

Figure 11: Minimum Function for N=10 M=6

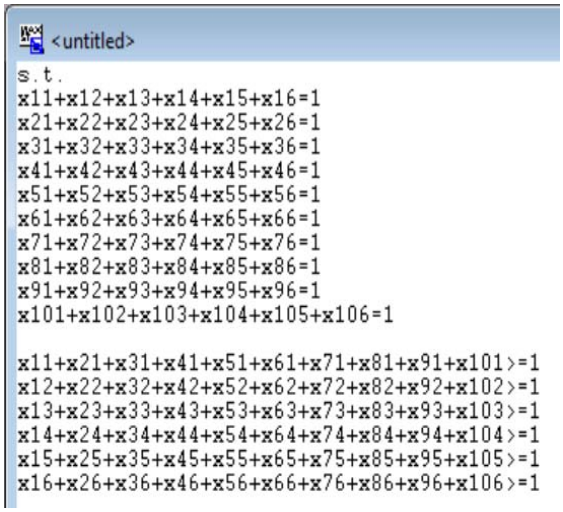

Figure 12: Constraint for N=10 M=6

The X and Y values, it is written in Figure 13 as follows:

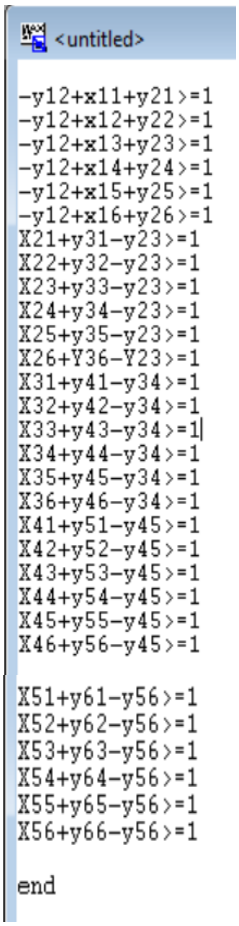

Figure 13: X and Y Values for N=10 M=6

The LINDO's execution result of the minimum function and constraints from the example above are shown in Table 4.

Table 4: Optimal Results for Variable X N=10 M=6

Variable	Value
X11	0,000000
X12	0,000000
X13	1,000000
X14	0,000000
X15	0,000000
X16	0,000000
X21	0,000000
X22	0,000000
X23	0,000000
X24	1,000000
X25	0,000000
X26	0,000000
X31	0,000000
X32	0,000000
X33	0,000000
X34	0,000000
X35	1,000000
X36	0,000000
X41	0,000000
X42	0,000000
X43	0,000000
X44	0,000000
X45	0,000000
X46	1,000000
X51	0,000000
X52	0,000000
X53	0,000000
X54	1,000000
X55	0,000000
X56	0,000000
X61	1,000000
X62	0,000000
X63	0,000000
X64	0,000000
X65	0,000000
X66	0,000000
X71	0,000000
X72	1,000000
X73	0,000000
X74	0,000000
X75	0,000000
X76	0,000000
X81	0,000000
X82	1,000000
X83	0,000000
X84	0,000000
X85	0,000000
X86	0,000000
X91	0,000000
X92	1,000000
X93	0,000000
X94	0,000000

X95	0,000000
X96	0,000000
X101	0,000000
X102	1,000000
X103	0,000000
X104	0,000000
X105	0,000000
X106	0,000000

4.4 Results Reading and Discussion

In accordance with the objectives of this study, objects are placed in clusters using a combinatorial approach based on mi function. the minimum specified. This new model forms a formula that shows if X = 1 then an object has precisely occupied a cluster optimally.

Based on Tables 2, 3, and 4, each variable X has a value of 1, then the first number after the variable is the i object that is placed in the j cluster. For example, X13 in Table 2, means that the 1st object occupies the 3rd cluster. X24 in Table 2 means the 2nd object occupies the 4th cluster, and so on.

For the X63 variable in Table 3, it means that the 6th object occupies the 3rd cluster. The variable X73 in Table 3 means that the 7th object occupies the 3rd cluster. Likewise, the variables contained in Table 4. The variable X54 in Table 4 means that the 5th object occupies the 4th cluster. The variables X72, X82, X92 and X102 in Table 4 are objects with the same cluster location, cluster 2. This shows that there are 4 objects in 1 cluster. The full test results on N = M and N> M are shown in Tables 5, 6 and 7.

Table 5: Cluster Optimization N=10 M=10

N Object	*Cluster*
1	3
2	4
3	5
4	6
5	7
6	8
7	9
8	2
9	10
10	1
1	9
6	9
2	1

Table 6: Cluster Optimization N=7 M=5

N Object	*Cluster*
1	3
2	4
3	5
4	1
5	2
6	3
7	3

Table 7: Cluster Optimization N=10 M=6

N Object	*Cluster*
1	3
2	4
3	5
4	6
5	4
6	1
7	2
8	2
9	2
10	2

The resulting of combinatorial optimization model shows test results that meet the desired requirements. This new model guarantees that no cluster is empty or has no members. Each cluster has at least 1 data or object as a member of the cluster. This condition can occur because of a function $\sum_{i=1}^{M} X_{ik} \geq 1, k = 1,...,M$. In addition, the combinatorial optimization model can also place multiple objects in the same cluster for both N = M and N> M.

Based on the explanation of the new approach produced, several differences can be seen from the existing cluster evaluation techniques, namely:

a. The combinatorial optimization approach uses the desired highest k test boundary so that direct data placement in clusters can be carried out.

b. Problem constraints are converted into combinatorial form.

c. The decision on the optimal cluster is obtained from X = 1, while X = 0 is not the optimal cluster.

In simple terms, the combinatorial optimization approach has formed an understandable algorithm, as can be seen in Table 8.

Table 8: Data Placement Algorithms in Clusters

1. Prepare the data along with the accompanying criteria and the highest cluster boundary
2. Determine the objective / objective function. This function is used to determine the final decision
3. Convert data to matrix form according to the amount of data and the highest cluster limitation
4. Perform the constraints formulation in the form of a non-linear formulation
5. Perform testing
6. If the test result for variable x is equal to 1, then the constant that follows x is the data that occupies the most optimal cluster. If the variable x is equal to 0, then the data and clusters are not optimal.

The algorithm formed is expected to guide data placement in clusters using the combinatorial optimization model. Based on the tests that have been carried out, the combinatorial optimization model is very likely to be an alternative in determining and placing data in clusters optimally.

5. CONCLUSION

Cluster optimization based on the combinatorial optimization approach has succeeded in providing an alternative in cluster evaluation techniques, especially in placing data or objects in clusters. The form of fast decisions through variables X and Y with values 0 and 1 can determine and place the data in the cluster optimally without having to test of k tests.

The resulting model shows the success of the test at N = M and N> M, where the number of objects is equal to the number of clusters and the number of objects is greater than the number of clusters. The linearization function is very helpful in placing objects or data in the cluster because it guarantees that each object or data has a similar closeness, is in the same cluster and each cluster has at least one object or data. In addition, the resulting model forms a solution path in the form of an algorithm that can be executed logically and easily.

In the application of clusters by users, empty clusters are avoided, because logically, grouping is made to show identity based on the data or objects in it. The user knows the identity of a cluster from the features or criteria attached to the objects that fill the cluster. This means that each cluster must provide useful information or knowledge. Therefore, this combinatorial approach is very helpful for users in overcoming these problems.

REFERENCES:

[1] C. D. Manning, P. Raghavan, and H. Schutze, *an Introduction To Informational Retreival*, no. April. 2009.

[2] Y. Zhang, T. Bouadi, and A. Martin, "An empirical study to determine the optimal k in Ek-NNclus method," in *5th International Conference on Beleif Functions (BELIEF2018)*, 2018, vol. 9, pp. 260–268, doi: 10.1007/978-3-319-99383-6_32.

[3] M. Henzinger, D. Leniowski, and C. Mathieu, "Dynamic Clustering to Minimize the Sum of Radii ∗," no. 48, pp. 1–10, 2017.

[4] T. Szkaliczki, "clustering.sc.dp: Optimal Clustering with Sequential Constraint by Using Dynamic Programming," *R J.*, vol. 8, no. 1, p. 318, 2019, doi: 10.32614/rj-2016-022.

[5] A. Lengyel and Z. Botta-Dukát, "Silhouette width using generalized mean—A flexible method for assessing clustering efficiency," *Ecol. Evol.*, vol. 9, no. 23, pp. 13231–13243, 2019, doi: 10.1002/ece3.5774.

[6] N. Kaoungku, K. Suksut, R. Chanklan, K. Kerdprasop, and N. Kerdprasop, "The silhouette width criterion for clustering and association mining to select image features," *Int. J. Mach. Learn. Comput.*, vol. 8, no. 1, pp. 69–73, 2018, doi: 10.18178/ijmlc.2018.8.1.665.

[7] M. R. Rao, "Cluster analysis and mathematical programming," *J. Am. Stat. Assoc.*, vol. 66, no. 335, pp. 622–626, 1971, doi: 10.1080/01621459.1971.10482319.

[8] S. S. Ghuman, "Clustering Techniques - A Review," *Int. J. Comput. Sci. Mob. Comput.*, vol. 5, no. 5, pp. 524–530, 2016, doi: 10.26438/ijcse/v6i6.10911099.

[9] I. H. Witten and E. Frank, *Data Mining:Practical Machine Learning Tools and Techniques second edition*. 2005.

[10] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 336, no., pp. 1–7, doi: 10.1088/1757-899X/336/1/012017.

[11] C. A. Sugar and G. M. James, "Finding the

Number of Clusters in a Dataset: An Information-Theoretic Approach," pp. 1–24, 2003, doi: 10.1198/016214503000000666.

[12] M. Villani *et al.*, "An Iterative Information-Theoretic Approach to the Detection of Structures in Complex Systems," *Complexity*, vol. 2018, no. 11, pp. 1–16, 2018, doi: 10.1155/2018/3687839.

[13] S. Xu, L. Zhang, P. Zhang, and H. Y. Noh, "An Information-Theoretic Approach for Indirect Train Traffic Monitoring Using Building Vibration," *Front. Built Environ.*, vol. 3, no. May, pp. 1–14, 2017, doi: 10.3389/fbuil.2017.00022.

[14] D. M. Budden and E. J. Crampin, "Information theoretic approaches for inference of biological networks from continuous-valued data," *BMC Syst. Biol.*, vol. 10, no. 89, pp. 1–7, 2016, doi: 10.1186/s12918-016-0331-y.

[15] M. Femy P.F and L. S. Mathew, "Clustering Dynamic and Distributed Dataset using Decentralized Algorithm," - *Int. J. Sci. Res. Dev.*, vol. 4, no. 04, pp. 1484–1486, 2016.

[16] A. Shafeeq B M and H. K S, "Dynamic Clustering of Data with Modified K-Means Algorithm," in *International Conference on Information and Computer Networks (ICICN 2012)*, 2012, vol. 27, pp. 221–225, doi: 10.13140/2.1.4972.3840.

[17] M. Elhoseny, K. Elleithy, H. Elminir, X. Yuan, and A. Riad, "Dynamic clustering of heterogeneous wireless sensor networks using a genetic algorithm, towards balancing energy exhaustion," *Int. J. Sci. Eng. Res.*, vol. 6, no. 8, pp. 1243–1252, 2015.

[18] P. . Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. 1987, pp. 53–65, 1987.

[19] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, 1979, doi: 10.1109/TPAMI.1979.4766909.

[20] T. Weise, *Global optimization algorithms: Theory and some applications*. 2008.

[21] M. Alameen, M. Abdul-Niby, T. Selmi, and S. Damrah, "A Combinatorial Optimization Approach to Solve the Synchronous Optical Network (SONET) Problem," *Glob. J. Enterp. Inf. Syst.*, vol. 6, no. 2, p. 4, 2014, doi: 10.15595/gjeis/2014/v6i2/51841.

[22] S. Dewi and Sawaluddin, "Combinatorial Optimization in Project Selection Using Genetic Algorithm," in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 300, no. 1, p. 6, doi: 10.1088/1757-899X/300/1/012020.

[23] R. L. Raschke, A. S. Krishen, P. Kachroo, and P. Maheshwari, "A combinatorial optimization based sample identification method for group comparisons," *J. Bus. Res.*, vol. 66, no. 9, pp. 1267–1271, 2013, doi: 10.1016/j.jbusres.2012.02.024.

[24] N. Huynh, J. Barthélemy, and P. Perez, "A heuristic combinatorial optimisation approach to synthesising a population for agent-based modelling purposes," *Jasss*, vol. 19, no. 4, p. 23, 2016, doi: 10.18564/jasss.3198.

[25] A. Shakir Hameed, B. Mohd Aboobaider, N. Hea Choon, M. Lafta Mutar, and W. Habib Bilal, "Review on the Methods to Solve Combinatorial Optimization Problems Particularly: Quadratic Assignment Model," *Int. J. Eng. Technol.*, vol. 7, no. 3.20, p. 15, 2018, doi: 10.14419/ijet.v7i3.20.18722.