Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14141

e-ISSN : 2541-2019
p-ISSN : 2541-044X

# A Mixed-Integer Programming Approach on Clustering Problems with Segmentation Application Customer

**Arin Elviana[1], Elly Rosmaini[2], Esther Sorta Mauli Nababan[3]***
[1] Magister of Mathematics, Universitas Sumatera Utara, Indonesia
[2][3] Department of Mathematics, Universitas Sumatera Utara, Indonesia
[1]elvianaarin@gmail.com, [2]elly1@usu.ac.id, [3]esther@usu.ac.id

**Submitted** :Sept 29, 2024 | **Accepted** : Oct 1, 2024 | **Published** : Oct 30, 2024

**Abstract:** As a marketing strategy, segmentation involves categorizing customers into specific groups based on their loyalty to a brand. This process is crucial in shaping an effective business strategy, as identifying various customer types enables businesses to target their marketing efforts more precisely. This research focuses on solving the cluster optimization problem by applying a combinatorial optimization approach to develop a cluster optimization method. The combinatorial optimization utilized here operates on a binary system, using 0s and 1s to identify the optimal cluster for each object. Specifically, a value of 1 indicates that an object is assigned to an optimal cluster, while a value of 0 signifies that the object belongs to a non-optimal cluster. By designating clusters with a value of 1, the method ensures that the best optimization value is achieved. The 0-1 non-linear problem model ensures that objects with the shortest distances between them are grouped in the same cluster. Additionally, the model guarantees that each object belongs to only one cluster and that, across k tests, every cluster contains at least one object. This model can also be used to determine the ideal number of clusters for a given dataset, ensuring optimal segmentation results for business applications.

**Keywords:** Clustering; Combinatorial Optimization Approach; Mixed Integer Programming; Marketing strategy; Segmentation

## INTRODUCTION

As a marketing tactic, segmentation entails grouping clients into subsets defined by their level of devotion to a brand. One of the first stages in developing a company strategy is identifying different types of customers. The marketing notion of segmentation remains significant, even within the framework of relationship marketing. Building stronger connections with consumers is gaining popularity since it leads to a clearer picture of what those customers want (Adiana et al., 2018). This understanding is crucial in today's competitive marketplace, where tailored marketing strategies can significantly enhance customer satisfaction and loyalty.

Finding out how customers behave and then using that information to develop an effective marketing plan is what customer segmentation is all about. By analyzing customer data, businesses can identify distinct segments characterized by similar behaviors, preferences, and needs. Each group already has commonalities in terms of behavior and requirements, allowing for a more targeted marketing process—including communication, products/services, and activities. This targeted approach not only optimizes marketing resources but also increases the likelihood of successful customer engagement and conversion (Khoa, 2020; Suryawijaya & Wardhani, 2023).

Data mining methods are used to accomplish grouping. In data mining, clustering is a crucial tool for dividing data into many groups defined by shared characteristics (Hananto et al., 2017). Clustering techniques can uncover hidden patterns within large datasets, enabling marketers to understand their customer base more comprehensively. By employing sophisticated algorithms, businesses can segment customers more accurately and dynamically adapt their strategies in response to evolving consumer trends.

One innovative approach in this realm is the application of Mixed Integer Programming (MIP), which serves as a powerful model for partitioning data sets into clusters. This optimization technique allows marketers to refine their segmentation processes by minimizing the maximum diameter of the resulting clusters, thereby achieving greater compactness and cohesion within each segment. The goal is to ensure that each customer group is not only distinct but also internally homogeneous, which enhances the effectiveness of targeted marketing strategies.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14141
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In summary, effective customer segmentation is essential for building lasting relationships with clients and optimizing marketing efforts. By leveraging advanced data mining techniques and optimization models like Mixed Integer Programming, businesses can gain deeper insights into their customer base, leading to more strategic decision-making and improved overall performance in the marketplace.

## LITERATURE REVIEW

The goal of clustering is to maximize the degree of similarity between items inside a given cluster, which is a method for organizing data that describes the relationships between various things (Santoso, 2010).

Clustering is a technique for discovering and organizing data sets that have commonalities between them (Jollyta et al., 2021). According to (Alfina et al., 2012), data mining techniques include clustering. If you're doing data mining and looking for distribution patterns in your dataset, clustering is a great tool to employ. The proximity of the attribute values used to characterize data objects—which are often shown as points in space—is typically used to derive object similarity. The Hierarchical and Partitioning approaches are two of several available options for clustering (Sembiring Brahmana et al., 2020; Shetty & Singh, 2021). Data is organized into a cluster tree or hierarchy using hierarchical techniques, and data items are directly clustered into many clusters using partitioning methods. Choosing the target number of clusters (two, three, etc.) is the first step in the partitioning clustering approach. Without first establishing a hierarchical process, the cluster process is executed once the number of clusters is known. A popular name for this technique is K-Means Clustering (Santoso, 2010).

There is a subfield of optimization known as combinatorial optimization. Finding the optimal solution to an optimization issue in a discrete or reducible collection of possible options is the subject of combinatorial optimization. Optimization in combinatorial systems is a subfield of combinatorial science and applied mathematics that has many connections to theoretical and practical areas such as computational complexity, algorithm theory, and applied research.

Algorithms and approaches are used in tandem with combinatorial optimization. Several domains, including data mining, may benefit from this method's ability to address specific issues. Finding the shortest possible network size while still meeting the capacity needs of Synchronous Optical Network (SONET) consumers is tackled using combinatorial optimization in (Alameen et al., 2014). The Capacitated Vehicle Routing Problem (CVRP) was analyzed to acquire the results. To further streamline the project selection process (Dewi & Sawaluddin, 2018), decrease the number of variables, and save money on survey sample space (Raschke et al., 2013), combinatorial optimization using Genetic Algorithms is a good idea. Clustering census data according to six demographic characteristics is another area that may be optimized via combinatorial optimization (Huynh et al., 2016).

According to the research, cluster optimization methods for resolving clustering issues prioritize using the closest distance between cluster members ($N$ objects) and a fixed number of $k$ tests. Using a combinatorial technique, as described in (Rao, 1971), one may place $N$ items into a number of clusters and optimize the number of objects in each cluster. The number of clusters depends on the application being utilized. A novel method for optimally clustering data or objects in a combinatorial form is the focus of the present researcher's study. As illustrated by the number $N$ of items that may be created directly and fill the cluster ideally, the combinatorial solution is apparent. In this combinatorial issue, the best cluster criteria are cluster distance and data placement inside clusters utilizing intra clusters. To accomplish the merging, the conditions of a clustering issue are taken into account. By combining the numbers 0 and 1, we can display the ideal cluster arrangement of $N$ items; the objects that were chosen for the cluster will be represented by 1, while the ones that were not chosen will be represented by 0.

## METHOD

The stage taken in this study to designing the model are as follows:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14141
e-ISSN : 2541-2019
p-ISSN : 2541-044X

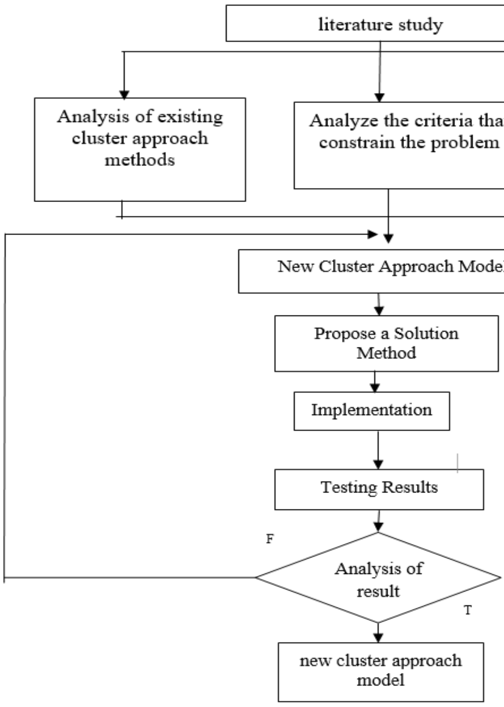


Fig. 1. Block Diagram of Research Procedure

Finding the ideal number of clusters is challenging since cluster appearance is a factor. Under these circumstances, similarity forms the basis of the problem criteria that transform into the constraint function. By measuring the distance between items, we may determine how similar they are. When we minimize the greatest distance inside a set of objects or data, we get the distance in question. Here is how data grouping is mimicked:

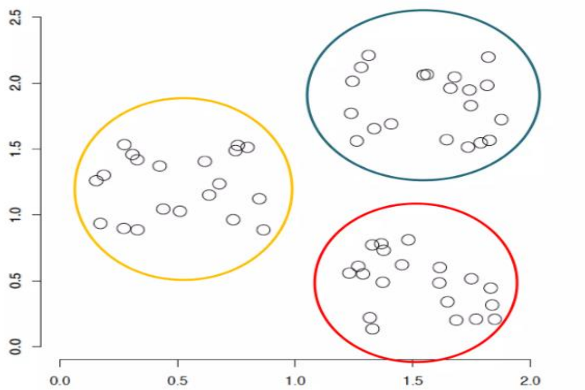


Fig.2. Illustration of Grouping Objects or Data

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

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

With constraints:

$$\sum_{k=1}^{M} X_{ik} = 1\,;\, i = 1\,,\dots\ \,,N$$

$$\sum_{i=1}^{M} X_{ik} \geq 1\,;\, k = 1\,,\dots\ \,,M$$

$$X_{ik} \in \{0,1\}\,;\, i = 1, \dots, N\,;\, k = 1, \dots, M$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14141
e-ISSN : 2541-2019
p-ISSN : 2541-044X

## RESULT

The number of clusters ($k$) must be determined in advance. The assignment of data to a cluster is indicated by $X = 1$. In most cluster evaluation techniques, such as the Elbow method and the Davies-Bouldin Index (DBI), the optimal number of clusters is obtained through the testing of various $k$ values, automatically optimizing the data within each cluster (Enza Wella et al., 2023). However, in this combinatorial model, testing for $k$ is unnecessary. Instead, the initial determination of $k$ serves as the upper limit for subsequent tests. The constant $i$ represents the optimal object that will be assigned to the optimal cluster $j$ indicated by $X = 1$.

### $N = M$ Combinatorial Optimization Approach

To construct the desired combinatorial optimization equation, let us consider five objects denoted by $N$ and five clusters represented by $M$. The coefficients of all $Y$ variables represent the distances among the objects. The resulting minimum objective function is expressed as follows:

$$\text{Minimum}: 10y_{12} + 15y_{13} + 9y_{14} + 16y_{15} + 20y_{23} + 19y_{24} + 16y_{25} + 16y_{34} + 13y_{35} + 15y_{45} + 10y_{21} + 15y_{31} + 9y_{41} + 16y_{51} + 20y_{32} + 19y_{42} + 16y_{52} + 16y_{43} + 13y_{53} + 15y_{54}$$

The constraints of the optimization problem arise from the necessity of having a defined number of clusters. These constraints are articulated as follows:

$$\begin{aligned}
x_{11} + x_{12} + x_{13} + x_{14} + x_{15} &= 1 \\
x_{21} + x_{22} + x_{23} + x_{24} + x_{25} &= 1 \\
x_{31} + x_{32} + x_{33} + x_{34} + x_{35} &= 1 \\
x_{41} + x_{42} + x_{43} + x_{44} + x_{45} &= 1 \\
x_{51} + x_{52} + x_{53} + x_{54} + x_{55} &= 1 \\
x_{11} + x_{21} + x_{31} + x_{41} + x_{51} &\geq 1 \\
x_{12} + x_{22} + x_{32} + x_{42} + x_{52} &\geq 1 \\
x_{13} + x_{23} + x_{33} + x_{43} + x_{53} &\geq 1 \\
x_{14} + x_{24} + x_{34} + x_{44} + x_{54} &\geq 1 \\
x_{15} + x_{25} + x_{35} + x_{45} + x_{55} &\geq 1
\end{aligned}$$

The relationships between $X$ and $Y$ values are formulated as follows:

$$\begin{aligned}
-y_{12} + x_{11} + y_{21} &\geq 1 \\
-y_{12} + x_{12} + y_{22} &\geq 1 \\
-y_{12} + x_{13} + y_{23} &\geq 1 \\
-y_{12} + x_{14} + y_{24} &\geq 1 \\
-y_{12} + x_{15} + y_{25} &\geq 1 \\
x_{21} + y_{31} - y_{23} &\geq 1 \\
x_{22} + y_{32} - y_{23} &\geq 1 \\
x_{23} + y_{33} - y_{23} &\geq 1 \\
x_{24} + y_{34} - y_{23} &\geq 1 \\
x_{25} + y_{35} - y_{23} &\geq 1 \\
x_{31} + y_{41} - y_{34} &\geq 1 \\
x_{32} + y_{42} - y_{34} &\geq 1 \\
x_{33} + y_{43} - y_{34} &\geq 1 \\
x_{34} + y_{44} - y_{34} &\geq 1 \\
x_{35} + y_{45} - y_{34} &\geq 1 \\
x_{41} + y_{51} - y_{45} &\geq 1 \\
x_{42} + y_{52} - y_{45} &\geq 1 \\
x_{43} + y_{53} - y_{45} &\geq 1 \\
x_{44} + y_{54} - y_{45} &\geq 1 \\
x_{45} + y_{55} - y_{45} &\geq 1
\end{aligned}$$

End

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

Using the LINDO application, the minimum function and constraints are then evaluated, yielding the following results:

*name of corresponding author

DOI : https://doi.org/10.33395/sinkron.v8i4.14141
e-ISSN : 2541-2019
p-ISSN : 2541-044X

$$X_{13} = 1; X_{24} = 1;\ X_{35} = 1; X_{42} = 1; X_{51} = 1$$

This outcome indicates that object 1 is placed in cluster 3, object 2 in cluster 4, object 3 in cluster 5, object 4 in cluster 2, and object 5 in cluster 1, achieving a minimum value of 168.

**$N > M$ Combinatorial Optimization Approach**

The subsequent analysis involves scenarios where the number of objects exceeds the number of clusters, denoted as $N > M$. The combinatorial model is tested to ascertain whether it can accurately determine the optimal cluster and allocate the best object to the most suitable cluster. The model will be simulated with $N = 7$ and $M = 5$. The minimum function for $N = 7$ and $M = 5$ is depicted in the following figure:

```
Minimum
10y12+15y13+9y14+16y15
+20y23+19y24+16y25
+16y34+13y35
+15y45+

10y21+15y31+9y41+16y51
+20y32+19y42+16y52
+16y43+13y53
+15y54
```

Fig.3. Minimum Function for $N = 7$, $M = 5$

Following the same methodology as the $N = M$ combinatorial optimization approach, the LINDO application is employed to evaluate the minimum function and constraints, resulting in the following findings:

$$X_{13} = 1; X_{24} = 1;\ X_{35} = 1; X_{41} = 1; X_{52} = 1; X_{63} = 1; X_{73} = 1$$

This outcome signifies that object 1 resides in cluster 3, object 2 in cluster 4, object 3 in cluster 5, object 4 in cluster 1, object 5 in cluster 2, and both objects 6 and 7 in cluster 3, with a minimum value of 168.

## DISCUSSIONS

The preceding description indicates that the initial step is to determine the number of clusters, denoted as $k$. The variable $X = 1$ dictates the allocation of data points within a cluster. Generally, the data included in the optimal cluster, derived from testing various values of $k$, is considered optimal; this holds true for most cluster evaluation methods, including the Elbow method and the Davies-Bouldin Index (DBI). However, in this combinatorial model, it is unnecessary to evaluate $k$. Instead, the initial value of $k$ serves as the upper bound for all subsequent tests. When $X = 1$, the ideal cluster $j$ is identified, and the constant iii represents the object to be assigned to that cluster. This approach can generate an optimal cluster size that is directly proportional to the number of objects. Furthermore, employing a combinatorial optimization method often results in fewer clusters than the actual number of items.

## CONCLUSION

Based on the results of the discussion on cluster optimization with the acquisition of the optimal number of clusters in combinatorial problems determined by the cluster distance. In addition, the optimal number of clusters can be influenced by several factors such as the data set for testing, the number of constraints, the calculation of similarity or dissimilarity based on distance, the size scale used and the clustering algorithm used and produces a decision variable in combinatorial form with members 0 and 1. The result of this model is to minimize the distance between objects in the same cluster, taking into account several constraints.

## REFERENCES

Adiana, B. E., Soesanti, I., & Permanasari, A. E. (2018). Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering. *Jurnal Terapan Teknologi Informasi*, *2*(1), 23–32. https://doi.org/10.21460/jutei.2018.21.76

Alameen, M., Abdul-Niby, M., Selmi, T., & Damrah, S. (2014). A Combinatorial Optimization Approach to Solve the Synchronous Optical Network (SONET) Problem. *Global Journal of Enterprise Information System*, *6*(2), 9. https://doi.org/10.15595/gjeis/2014/v6i2/51841

Alfina, T., Santosa, B., & Barakbah, A. R. (2012). Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Keduanya dalam Cluster Data (Studi Kasus: Problem Kerja Praktek Teknik Industri). *Jurnal Teknik ITS*, *1*(1), A521–A525.

Dewi, S., & Sawaluddin. (2018). Combinatorial Optimization in Project Selection Using Genetic Algorithm. *IOP Conference Series: Materials Science and Engineering*, *300*, 012020. https://doi.org/10.1088/1757-899X/300/1/012020

Enza Wella, Y., Okfalisa, O., Insani, F., Saeed, F., & Che Hussin, A. R. (2023). Service quality dealer

*name of corresponding author

identification: the optimization of K-Means clustering. *SINERGI*, *27*(3), 433. https://doi.org/10.22441/sinergi.2023.3.014

Hananto, V. R., Churniawan, A. D., & Wardhanie, A. P. (2017). Perancangan Analytical CRM untuk Mendukung Segmentasi Pelanggan di Institusi Pendidikan. *Jurnal Ilmiah Teknologi Informasi Asia*, *11*(1), 79. https://doi.org/10.32815/jitika.v11i1.55

Huynh, N., Barthelemy, J., & Perez, P. (2016). A Heuristic Combinatorial Optimisation Approach to Synthesising a Population for Agent Based Modelling Purposes. *Journal of Artificial Societies and Social Simulation*, *19*(4). https://doi.org/10.18564/jasss.3198

Jollyta, D., Efendi, S., Zarlis, M., & Mawengkang, H. (2021). A Combinatorial Optimization Approach to Determining Optimal Data in Cluster. *2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, 1–5. https://doi.org/10.1109/AIMS52415.2021.9466087

Khoa, B. T. (2020). The Antecedents of Relationship Marketing and Customer Loyalty: A Case of the Designed Fashion Product. *The Journal of Asian Finance, Economics and Business*, *7*(2), 195–204. https://doi.org/10.13106/jafeb.2020.vol7.no2.195

Rao, C. R. (1971). Unified Theory of Linear Estimation. *Sankhyā: The Indian Journal of Statistics, Series A*, *33*(4), 371–394.

Raschke, R. L., Krishen, A. S., Kachroo, P., & Maheshwari, P. (2013). A combinatorial optimization based sample identification method for group comparisons. *Journal of Business Research*, *66*(9), 1267–1271. https://doi.org/10.1016/j.jbusres.2012.02.024

Santoso, S. (2010). *Statistik Multivariat*. Elex Media Komputindo.

Sembiring Brahmana, R. W., Mohammed, F. A., & Chairuang, K. (2020). Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, *11*(1), 32. https://doi.org/10.24843/LKJITI.2020.v11.i01.p04

Shetty, P., & Singh, S. (2021). Hierarchical Clustering: A Survey. *International Journal of Applied Research*, *7*(4), 178–181. https://doi.org/10.22271/allresearch.2021.v7.i4c.8484

Suryawijaya, T. W. E., & Wardhani, M. F. (2023). Tailoring the future of MSME marketing: A study on leveraging customer data for personalized experiences. *Implementasi Manajemen & Kewirausahaan*, *3*(1), 76–88. https://doi.org/10.38156/imka.v3i1.163