Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754
P-ISSN : 2580-569X
Vol. 8, No. 2, 2024
DOI: 10.30829/zero.v8i2.21581
pp. 65-71

Cluster Analysis of Factors Affecting the Amount of Shipping Costs PT. Indah Logistik Cargo Lampung Branch Using Average Linkage Method

Ranara Athalla Yoka[1], Bernadhita Herindri Samodera Utami[2], Riza Sawitri[3]

[1,2,3]Department of Mathematics, Universitas Lampung, Bandar Lampung, Indonesia

Article Info

Article history:

Received, 09 10 2024
Revised, 25 10 2024
Accepted, 16 12 2024

Keywords:

Cluster Analysis
Average Linkage Method
PT. Indah Logistic Cargo

ABSTRACT

Cluster analysis analyzes similar elements as different and independent cluster research objects (not interconnected). PT. Indah Logistik Cargo is a company that provides goods and motorcycle shipping services through three channels: land, sea, and air. This research has three objectives. The first is to categorize postage prices based on the type of goods and services provided so that customers can easily see the options available and understand the differences between each option. Thirdly, categorize prices so companies can offer special promotions or discounts for certain services. Fourth, cluster prices so companies can offer special promotions or discounts for certain services. Based on the analysis results, 3 cluster groups were formed: cheap, medium, and expensive. It can be concluded that the group with low postage prices (Rp 16,160 to Rp 622,160) consists of Medan, Bengkulu, Palembang, Padang, and Pekanbaru; the group with medium postage prices (Rp 40,400 to Rp 888,800) is only Banda Aceh; and the group with high postage prices (Rp 62,620 to Rp 1,555,400) consists of Tanjung Pinang and Pangkal Pinang.

Corresponding Author:

Bernadhita Herindri Samodera Utami,
Department of Mathematics,
Universitas Lampung, Bandar Lampung, Indonesia
Email: ind.indri1245@gmail.com

1. INTRODUCTION

Cluster analysis groups similar elements as research objects into different and independent clusters (not interconnected). This is in contrast to discriminant analysis, where the cluster has already been determined, and the discriminant function is used to determine which cluster an element or object should be in [1].

Cluster analysis is a multivariate technique, but the concept of variate in this technique differs from that concept of variate in other multivariate methods. In other techniques, variates are defined as linear combinations of variables. In contrast, in cluster analysis, variates are defined as several of variables (considered as characteristics) that compare an object with other objects [2]. So, in cluster analysis, there is no empirical search for variate values as in other multivariate techniques. The principal aim of cluster analysis is to categorize a collection of items into two or more clusters based on their similarity across numerous attributes [3,4].

PT Indah Logistik Cargo is a company that provides shipping services for goods and motorbikes through three channels: land, sea, and air. To find out Indah Logistik Cargo's shipping rates, people can visit the official website at https://indahonline.com/en, enter the city of origin and destination address, and

type in the security code. More information about shipping costs will be displayed based on the route and type of service selected. The grouping of shipping prices aims to make it easier for customers to understand and choose services according to their needs.

This research has three objectives, first to categorize shipping prices based on the type of goods and services provided so that customers can easily see the options available and understand the differences between each option. Second, to categorise prices to ensure transparency in shipping costs. Third, to categorize prices so that companies can offer special promotions or discounts for certain services.

In logistics, cluster analysis is vital role in grouping shipping price data based on specific characteristics. One method used is the average linkage method. This method calculates the length of distance each pair of points in the dataset. It involves a complex distance matrix and calculating of the mean distance between all pairs of points in two clusters [5].

In the context of postage prices, this method helps logistics companies identify regions with similar pricing patterns. This information can also be used to optimize delivery routes, determine competitive prices, and set business strategies. Thus, cluster analysis of postage prices can help logistics companies or e-commerce businesses make decisions related to pricing, resource allocation, and delivery route planning [6].

2. RESEARCH METHOD

2.1 Cluster Analysis

Cluster analysis is a multivariate technique employed to categorize items according to their similarities. Objects that are relatively close to one another are categorized into distinct groups [7].

The benefits of clustering include organizing large sets of observational data with numerous variables, simplifying data analysis by reducing it into clusters, and its applicability to ordinal, interval, and ratio scales [5].

The drawbacks of cluster analysis include the potential for researcher subjectivity, as it often relies solely on interpreting dendrogram images. In cases of heterogeneous data across research objects, it can be challenging to determine the appropriate number of groups. Different methods can yield significantly varying results, necessitating comparisons between techniques. Additionally, as the number of observations increases, the likelihood of errors also rises [2].

The cluster formation procedure is categorized into two categories: hierarchical and non-hierarchical. Hierarchical clustering creates a hierarchical or tree-like branching structure. In this method, the number of groups to be formed is not predetermined, and clustering is performed through either sequential merging (agglomerative) or division (divisive). In contrast, the non-hierarchical method requires specifying the desired number of clusters beforehand, and the clustering process is carried out without adhering to a hierarchical structure [5].

2.2 Hierarichal Clustering Method

This method is used to cluster observations in a structured manner based on their similar characteristics. The number of desired groups is unknown. This method is divided into two hierarchical methods: the agglomerative method and the divisive method.

The agglomerative method begins with the assumption that every object represents an own cluster, it should be noted that merging in this method always involves a distance matrix, and then the results of group analysis from this method can be presented as a dendrogram. The divisive method is a clustering process with a top-down approach, starting with N objects grouped into one group. Then the group is partitioned into two groups at each step until N groups are obtained, with each group having one object. The basis of clustering is also based on distance. However, this technique is not widely used, so few procedures have been developed [8].

The agglomerative technique includes three main methods: the linkage method, the variance method, and the centroid method. The linkage method is categorized into single linkage, complete linkage, and average linkage methods, while the variance method includes the Ward method.

The single linkage method, also known as the nearest neighbor method, calculates the distance between two clusters based on the shortest (smallest) distance between any member of one cluster and any member of the other cluster. The complete linkage method, or furthest neighbor method, determines the distance between two clusters using the longest (largest) distance between any member of one cluster and any member of the other.

The average linkage method determines the distance between clusters by averaging all the pairwise distances between members of the two clusters. The Ward method minimizes the increase in total within-cluster variance when merging two clusters [6]. According to [9], the hierarchical clustering approach operates by utilizing a set of N items to be clustered with a N×N matrix that represents the distances between these items:

1. Start by creating N clusters; each cluster has an item. Suppose the distance between clusters equalsthe distance between the items they contain.
2. Identify the closest pair of clusters, and create a new cluster. Thus, we currently possess N - 1 clusters.
3. Determine the distance between the new cluster and each of the others.
4. Continue doing steps 2 and 3 until all items cluster with N items. Clearly, consolidating N things a single large cluster is futile.

2.3 Average Linkage Method

The average linkage method is a clustering technique that calculates the mean distance between all possible pairs of objects inside one cluster and all objects in another cluster. The method determines the distance between two clusters by averaging the distances calculated within each cluster. The procedure for average linkage begins by defining D = $\{d_{ik}\}$ to obtain the closest objects, for example, U and V. Then this object is combined into a cluster (UV), and then the distance between (UV) and other clusters (W), [10].

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{UV} N_W} \quad (1)$$

where:

d_{ik} = distance between object *i* in *cluster* (*UV*) and object *k* in *cluster W*

N_{UV} = number of items in *cluster UV*

N_W = number of items in *cluster W*

2.4 Distance Proximity Measure

Since cluster analysis seeks to identify vectors of similar observations and group them into clusters, many techniques use an index of similarity or closeness between each pair of observations. A commonly used proximity approach is to measures the similarity expressed in terms of the distance between pairs of objects. The smaller the magnitude of the distance of an individual to another individual, the more similar the individuals are so that the individuals will be included in the same group [11].

In cluster analysis, Euclidean distance is used as a proximity measure [12, 13, 14], which is defined as follows:

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(\boldsymbol{x}-\boldsymbol{y})'(\boldsymbol{x}-\boldsymbol{y})} = \sqrt{\sum_{j=1}^{p}(x_j - y_j)^2} \text{ or } d(i, j) = \sqrt{\sum_{j=1}^{p}(x_{ik} - x_{jk})^2} \quad (2)$$

where:

d_{ij} = the distance between *i*-th object and *j*-th object

p = number of *cluster* variables

x_{ik} = data from *i*-th subject on the *k*-th variable

x_{jk} = data from *j*-th subject on the *k*-th variable

3. RESULT AND ANALYSIS

The data used in this study include five factors that affect the amount of shipping costs of the PT. Indah Logistik Cargo Lampung branch in each of the nine cities in the Sumatra region in 2024. These factors include the price of shipping documents, packages (5-10kg), motorbikes (110cc), expedition distance, and expedition time.

Table 1. Data on The Total Shipping Cost of 9 Cities in The Sumatra Region Based on Document, Package, Motorbike, Expedition Distance, and Expedition Time

Destination	Document	5-10kg Package	110 cc Motorbike	Expedition Distance (Km)	Expedition Time (days)
Banda Aceh	40.400	80.800	888.800	2286	7
Medan	28.280	56.560	622.160	1678	4
Bengkulu	20.200	40.400	444.400	598	3

Palembang	16.160	32.320	355.520	336	2
Padang	26.260	52.520	577.720	1172	3
Pekanbaru	20.200	40.400	444.400	1051	3
Jambi	16.160	32.320	355.520	584	3
Tanjung Pinang	70.700	141.400	1.555.400	1011	12
Pangkal Pinang	62.620	125.240	1.377.640	582	10

Before conducting the analysis, standardization is necessary prior to doing the study if there is variability in units. Table 1, we will standardize the data to z-score. By using SPSS software, the z-score value is obtained as follows:

Table 2. Data on the Total Postage of 9 Cities in Sumatra Region Based on Document, Package, Motorbike, Expedition Distance, and Expedition Time in the Form of *z-Score*

Destination	Document	5-10kg Package	110 cc Motorbike	Expedition Distance (Km)	Expedition Time (days)
Banda Aceh	0,34184	0,34184	0,34184	2,01718	0,49412
Medan	-0,25363	-0,25363	-0,25363	1,03829	-0,33971
Bengkulu	-0,6506	-0,6506	-0,6506	-0,70054	-0,61765
Palembang	-0,84909	-0,84909	-0,84909	-1,12237	-0,8956
Padang	-0,35287	-0,35287	-0,35287	0,22361	-0,61765
Pekanbaru	-0,6506	-0,6506	-0,6506	0,0288	-0,61765
Jambi	-0,84909	-0,84909	-0,84909	-0,72308	-0,61765
Tanjung Pinang	1,83051	1,83051	1,83051	-0,0356	1,88385
Pangkal Pinang	1,43354	1,43354	1,43354	-0,7263	1,32796

Table 2 shows the factors affecting the amount of shipping in 9 cities in the Sumatra region that have been standardised to z-score. By using the data in Table 2, we will group the cities. By using SPSS software, the Euclidean distance is obtained as presented in Table 3 below:

Table 3. Euclidean Distance Matrix (*Stage* 1)

Case	Squared Euclidean Distance								
	1	2	3	4	5	6	7	8	9
1	0	2,717	11,577	16,043	5,901	8,145	13	12,794	11,797
2	2,717	0	3,574	6,041	0,77	1,569	4,243	19,128	14,434
3	11,577	3,674	0	0,373	1,12	0,532	0,119	25,168	16,817
4	16,043	6,041	0,373	0	2,628	1,521	0,237	30,447	20,732
5	5,901	0,77	1,12	2,628	0	0,304	1,635	20,626	14,261
6	8,145	1,569	0,532	1,521	0,304	0	0,684	24,73	17,387
7	13	4,243	0,119	0,237	1,635	0,684	0	28,271	19,417
8	12,794	19,128	25,168	30,447	20,626	24,73	28,271	0	1,259
9	11,797	14,434	16,817	20,732	14,261	17,387	19,417	1,259	0

Description:

Banda Aceh (1), Medan (2), Bengkulu (3), Palembang (4), Padang (5), Pekanbaru (6), Jambi (7), Tanjung Pinang (8), and Pangkal Pinang (9).

The results of the SPSS software analysis obtained the following cluster analysis results:

Table 4. Agglomeration Schedule using Between Group Linkage Method

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	7	0,119	0	0	3
2	5	6	0,304	0	0	4
3	3	4	0,305	1	0	6
4	2	5	1,170	0	2	6
5	8	9	1,259	0	0	8
6	2	3	2,442	4	3	7
7	1	2	9,564	0	6	8
8	1	8	19,715	7	5	0

Table 4 shows the agglomeration schedule with the 'between group linkage' method. After measuring the Euclidean distance between individual variables is measured, clustering is done in stages.

Stage 1: A cluster is established, comprising of samples 3 (Bengkulu) and 7 (Jambi) with a distance of 0.119

Stage 2: A cluster is established, comprising of samples 5 (Padang) and 6 (Pekanbaru) at a distance of 0,304.

The agglomeration process begins by merging the two closest objects, identified by the shortest distance among the nine possible distance combinations. The progression of clustering is reflected in the final column, where the number 3 indicates that the next clustering step occurs at stage 3. From there, the process continues sequentially through stages, such as moving from stage 3 to stage 6, and so on, until the final stage is reached.

Cluster membership can be interpreted as the degree of association or fusion of a data point with a particular cluster, which is often measured in numerical form.

The following is the cluster membership output with the help of SPSS software:

Table 5. Cluster Membership by Software SPSS

Case	3 Clusters
1: Banda Aceh	1
2: Medan	2
3: Bengkulu	2
4: Palembang	2
5: Padang	2
6: Pekanbaru	2
7: Jambi	2
8: Tanjung Pinang	3
9: Pangkal Pinang	3

In Table 5, the results show that 3 clusters (cheap, medium, and expensive groups) will be formed with the following members.

Cluster 1 (Group 1): Banda Aceh

Cluster 2 (Group 2): Medan, Bengkulu, Palembang, Padang, Pekanbaru, Jambi

Cluster 3 (Group 3): Tanjung Pinang, Pangkal Pinang

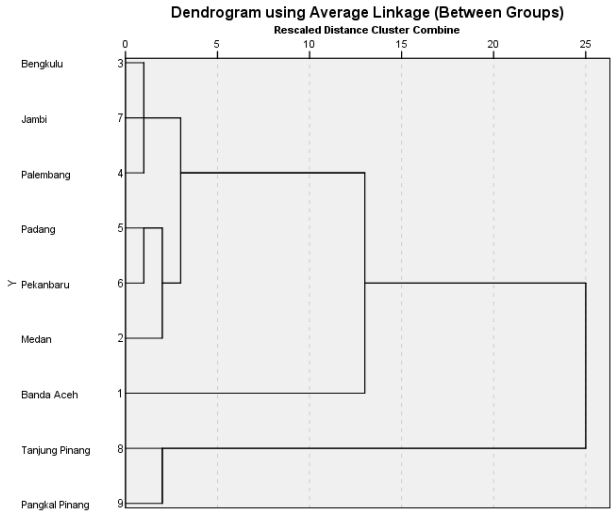

Figure 1. Dendogram

Dendrograms are visual representations of the steps in a clustering analysis solution that show how clusters are formed and the distance coefficient values at each step. In the dendrogram, the actual distances are rescaled in values from 0 to 25 with a constant ratio of distances between steps. It can be seen that the researcher wants to create 3 clusters as seen in the dendrogram, so cluster 1 consists only of sample 1 (Banda Aceh); cluster 2 consists of sample 2 (Medan), sample 3 (Bengkulu), sample 4 (Palembang), sample 5 (Padang), sample 6 (Pekanbaru); and cluster 3 consists of sample 8 (Tanjung Pinang) and sample 9 (Pangkal Pinang).

4. CONCLUSION

In light of the discourse presented in this study, it can be inferred that the results of the average linkage method produce three groups (clusters), namely groups with cheap shipping prices, groups with medium shipping prices, and groups with expensive shipping prices. After comparison, it can be concluded that the group members of each group are:

1) The groups with low postage prices (IDR 16,160 to IDR 622,160) are Medan, Bengkulu, Palembang, Padang, and Pekanbaru.
2) The only group with medium postage prices (IDR 40,400 to IDR 888,800) is the city of Banda Aceh.
3) Tanjung Pinang and Pangkal Pinang are the cities with the most expensive postage prices (IDR 62,620 to IDR 1,555,400).

REFERENCES

[1] M. G. H. Omran, A. P. Engelbrecht, and A. Salman, "An Overview of Clustering methods," *Intell. Data Anal.*, vol. 11, no. 6, pp. 583–605, 2007, doi: 10.3233/ida-2007-11602.

[2] M. Crum, T. Nelson, J. de Borst, and P. Byrnes, "The Use of Cluster Analysis in Entrepreneurship Research: Review of Past Research and Future Directions," *J. Small Bus. Manag.*, vol. 60, no. 4, pp. 961–1000, 2022, doi: 10.1080/00472778.2020.1748475.

[3] N. Sari, E. Sudriyanti, and N. Hanif, "Transport Network Planning for Freight Transport based on Environmental Approach," *Adv. Environ. Sci.*, vol. 13, no. 2, pp. 70–77, 2021, [Online]. Available: http://www.aes.bioflux.com.ro

[4] A. Saxena *et al.*, "A Review of Clustering Techniques and Developments," *Neurocomputing*, vol. 267, no. 1, pp. 664–681, 2017, doi: 10.1016/j.neucom.2017.06.053.

[5] S. Setyaningtyas, B. I. Nugroho, and A. Zaenul, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 10, no. 2, pp. 52–61, 2022, doi: 10.21063/jtif.2022.v10.2.52-61.

[6] P. Govender and V. Sivakumar, "Application of K-Means and Hierarchical Clustering Techniques for Analysis of Air Pollution: A review (1980–2019)," *Atmos. Pollut. Res.*, vol. 11, no. 1, pp. 40–56, 2020, doi: 10.1016/j.apr.2019.09.009.

[7] Supratno, J. (2004). *Analisis Multivariat Arti dan Interpretasi*. Jakarta: Rineka Cipta.

[8] Hardius Usman, dkk. (2013). *Aplikasi Teknik Multivariate untuk Riset Pemasaran.* Jakarta: PT Grafindo Persada. 8

[9] Hardle. W. dan L. Simar. (2007). *Applied Multivariate Statistical Analysis. 2nd Edition*. New York: Springer Berlin Heidelberg.

[10] Johnson dan Wichern. (2007). *Applied Multivariate Stastistical Analysis* 6^{th} *Edition*. New Jersey: Pretice-Hall, Inc. 10

[11] Johnson, S.C. (1967). *Hierarchical Clustering Schemes*. New Jersey: Pretice-Hall, Inc. 9

[12] Lind, Douglas A, dkk. (2007). *Teknik-teknik Statistika dalam Bisnis dan Ekonomi*. Jakarta: Salemba Empat.

[13] Narimawati, U. (2008). *Metodologi Penelitian Kulitatif dan Kuantitatif, Teori dan Aplikasi*. Bandung: Agung Media. 7

[14] Simamora, Bilson. (2005). *Analisis Multivariat Pemasaran Edisi Pertama*. Jakarta: PT Gramedia Pustaka Utama.