

Hybrid cluster analysis of customer segmentation of sea transportation users

Hybrid cluster
analysis

321

Bambang Eka Cahyana, Umar Nimran, Hamidah Nayati Utami and
Mohammad Iqbal

Faculty of Administrative Sciences, Universitas Brawijaya, Malang, Indonesia

Received 20 July 2019
Revised 10 August 2019
Accepted 2 September 2019

Abstract

Purpose – The purpose of this study is to apply hybrid cluster analysis in classifying PT Pelindo I customers based on the level of customer satisfaction with passenger services of PT Pelindo I.

Design/methodology/approach – Hybrid cluster analysis is a combination of hierarchical and non-hierarchical cluster analysis. This hybrid cluster analysis appears to optimize the advantages of hierarchical and non-hierarchical methods simultaneously to obtain optimal grouping. Hybrid cluster analysis itself has high flexibility because it can combine all hierarchical and non-hierarchical methods without any limits in the order of analysis used.

Findings – The results showed that 72% of PT Pelindo I customers felt PT Pelindo I service was special, while the remaining 28% felt PT Pelindo I service was good.

Originality/value – In total, 117 customers of PT Pelindo I were involved in a study using the non-probability sampling method.

Keywords Hybrid cluster analysis, PT pelindo I, Customer satisfaction, Customer segmentation, Sea transportation users

Paper type Research paper

1. Introduction

As a maritime country and the biggest archipelago country in the world, Indonesia possesses a vast sea area that offers a great number of resources. On one hand, marine resources are valuable assets that can be used for the prosperity of the nation. On the other hand, lands that are separated by sea become challenges for the citizen of Indonesia to travel across islands. As most of the Indonesian community belongs to the middle-low economic level, affordable sea transport modes are more preferred compared to air transport modes, which are relatively higher in price.

Responding to the high demand for sea transport, PT Pelindo I offers sea transportation services across islands including South Sumatera, Riau and Aceh. The company has 16 ports including passenger ports and container ports ranging from a prime class up to Class V. It is obvious that PT Pelindo I holds key roles to the economic and social activities among



© Bambang Eka Cahyana, Umar Nimran, Hamidah Nayati Utami and Mohammad Iqbal. Published in *Journal of Economics, Finance and Administrative Science*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) license. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this license may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Journal of Economics, Finance and
Administrative Science
Vol. 25 No. 50, 2020
pp. 321-337
Emerald Publishing Limited
2218-0648
DOI 10.1108/JEFAS-07-2019-0126

the surrounding society. As a service provider, it is important for PT Pelindo I to continue improving its service quality. In this research, the level of customer satisfaction toward the service of this company was measured using a set of questionnaires on customer satisfaction.

In addition, customer segmentation has been considered as the major focus in this research. Therefore, the segmentation analysis was also performed. This view goes in line with the theory of market segmentation regarding the importance of analyzing consumer characteristics (Kotler and Keller, 2013). Customer segmentation is necessary to be evaluated as results provide valuable insights for PT Pelindo I in making managerial decisions.

The characteristics of the population this research had to be comprehensively determined. As only a few research studies included only a single variable (Solimun, 2010), multivariate analysis was performed in analyzing the population characteristics in this research in the form of cluster analysis.

Cluster analysis is a multivariate statistical procedure that groups entities into relatively homogeneous groups (Latan, 2014). The segmentation facilitates the exploration and interpretation of the results of characteristic measurements. This develops the results of the questionnaire to become a representative reference in measuring the customer characteristics of PT Pelindo I.

Prior research has intrigued the researcher to apply hybrid cluster analysis in determining the customer segmentation of PT Pelindo I. This method allowed the researcher to simultaneously perform cluster hierarchy analysis and non-hierarchy analysis (Sukbukti, 2017). Hence, clusters can be determined from the results of the analysis based on customer characteristics. Further, the results of this analysis can be the benchmark of PT Pelindo I in improving its service quality to match the characteristics and needs of its customers.

2. Materials and methods

2.1 Cluster analysis

Cluster analysis is a statistical technique that groups research objects or variables in groups. Each object or variable is considered to have similarities in properties and characteristics (Hair *et al.*, 2010). Practically, cluster analysis is administered to segment some consumers (respondents) into several groups (clusters) based on the similarity of their attributes.

Cluster analysis in this research was done to group research objects based on their characteristic similarities. An ideal clustering offer:

- Intern homogeneity (within a cluster); similarities among members of a cluster.
- Extern homogeneity (between cluster); differences between clusters.

There are several methods of segmentation in cluster analysis as follows:

- *Hierarchy method*: segmentation starts from two or more objects that have the closest similarities before analyzing other objects to form a kind of “tree” that shows a clear level (hierarchy) between objects, from the ones that have close similarities to the least similar ones. A tool that is used in performing the hierarchical process is called “dendrograms.”
- *Non-hierarchy method*: starting from determining the intended number of clusters (two, three, etc). After the number of clusters is determined, the clustering process is carried out without following the hierarchy process. This method is commonly called the “K-means cluster.” K-means cluster is effectively and efficiently used to group many objects beyond 100 objects.

Hybrid method: hybrid method is a combination of hierarchical and non-hierarchical methods that extracts the advantages of both methods to determine the best cluster.

2.2 Hybrid cluster method

2.2.1 K-mean method. The materials K-mean cluster is a statistical analysis that is useful for grouping many objects into groups based on certain variables while background characteristics of the objects are not clearly defined (Hair *et al.*, 2010).

K-means is one of the clustering algorithms that divides research data into groups. This algorithm accepts input in data without class labels. The steps of the clustering algorithm are summarized as follows:

- Randomly determining the centroid based on the intended number of final clusters.
- Determining the clusters based on the closeness to the centroid.
- After grouping all objects into clusters, the gap between centroid to each cluster is measured.
- The measurement is repeated (iteration) from Steps 1-3 until the centroid of every cluster is convergent.

The clustering process in this research was conducted by measuring the closeness of certain data to its centroid point. Minkowski gap measurement could be performed to measure the gap between two data as follows:

$$d(x_i, x_j) = \left(|x_{i1} - x_{j1}|^g + |x_{i2} - x_{j2}|^g + \dots + |x_{ip} - x_{jp}|^g \right)^{1/g} \quad (1)$$

where:

$g = 1$, to measure the Manhattan gap;

$g = 2$, to measure the Euclidean gap;

$g = \infty$, to measure the Chebychev gap;

x_i, x_j = two data, which gap is to be measured; and

p = dimension of a data.

The centroid point can be renewed using the following formula.

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (2)$$

where:

μ_k = centroid point of cluster-K;

N_k = the number of data in cluster-K; and

x_q = data number- q in cluster-K.

2.2.2 Hierarchy method. Hierarchical clustering is a well-known clustering algorithm. The hierarchical clustering technique allows sequences of partitions to be made by:

- considering all research objects as an initial cluster;

- measuring the gap between the initial cluster;
- combining two initial clusters with the closest gap into one cluster; and
- repeating step number three until all objects are grouped into one single cluster.

The overall results of the hierarchical clustering algorithm are described in a tree-shaped graphic called a dendrogram. Dendrogram is obtained by combining the second line clusters with the clusters at the closest distance to form a single cluster. Meanwhile, in this research, the average linkage method was performed.

Average linkage method considers the distance between two clusters as the average distance among all members in a cluster and the members of other clusters based on this following formula (Hair *et al.*, 2010):

$$d_{(I)K} = \frac{\sum \sum d_{ab}}{N_I N_K} \quad (3)$$

where:

d_{ab} = gap between objects in cluster-I and objects in cluster-K;

N_I = number of items in cluster-I; and

N_K = number of items in cluster-I and K.

2.2.3 Hybrid method. The hybrid method was first introduced by Chipman and Tibshirani (2005). This method takes advantage of both hierarchical and non-hierarchical methods in determining the clusters. This method was redeveloped by comparing several possible hybrid methods (Cheu *et al.*, 2004).

The hybrid method used in this study is the combination of the average linkage and non-hierarchical methods (K-means) through these following steps:

- Performing an average linkage algorithm to obtain cluster-k.
- Calculating the central value of each cluster formed.
- Conducting the K-means algorithm based on the results of Step 2.

The above measurement resulted in the combination of those two clustering methods. The coding of the hybrid cluster method is presented in Attachment 4.

2.2.4 Euclidian gap measurement. The Euclidean method measures the values of two variables. This method is considered easy and it takes a shorter time as its process is simple. Euclidean is a heuristic function that is obtained based on a barrier-free gap to obtain the value from the length of the diagonal line in a triangle.

However, before obtaining those values, both centroid points should be put into two-dimensional coordinates (x, y). Two p_1 points = (x_1, y_1) and $p_2 = (x_2, y_2)$ are put into this following equation (Euclidian formula):

$$d = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \quad (4)$$

2.2.5 Market segmentation. Market segmentation is the grouping of markets into homogeneous consumer groups, where each group (part) can be chosen as a target market (targeted) for the marketing of a product.

For market segmentation or market grouping to run effectively, these following market grouping requirements should be fulfilled.

- Measurability, characteristics of the properties of the customers are measurable and observable.
- Accessibility, a condition where a company effectively directs its marketing to a certain targeted market segment.
- Sustainability, the market segment should be large and profitable to be targeted.

While, effective market segmentation is characterized by these points (Ciptono, 2001):

- *Measurable*: including its scope, purchasing power and segment profile.
- *Size of the segment (substantial)*: the size should be large enough and profitable to be served.
- *Accessible*: the market can be effectively approached.
- *Differentiable*: the market is conceptually distinct and it gives different responses toward the elements of a certain innovative program.

Actionable: effective marketing programs can be designed to attract the market segment.

2.2.6 *Types of market segmentation*. Market segmentation is classified into several types including Kotler and Keller (2013):

- *Geographic-based market segmentation*
This segmentation groups the market based on geographic areas such as countries, regencies, cities and villages. Geographic areas that are considered profitable will be set as the target of company operation.
- *Demographic-based market segmentation*
This segmentation divides the market into groups based on age, sex, economic level and educational background.
- *Psychographic-based market segmentation*
This segmentation segments the market based on lifestyle, social class and personalities.

According to Ciptono (2001), effective market segmentation is as follows.

- Measurable, size, purchasing power and segment profile.
- Substantial: large enough and profitable to be served.
- Accessible: actively accessible and served.
- Differentiable: conceptually can be separated and give different responses to the elements and mix programs.
- Actionable: effective programs can be formulated to attract and serve the segment.

2.2.7 *Research variables*. A variable is an object that is being observed in research. It is often regarded as the most influential factor in research or phenomena that is being researched. Research variables are the construct or characteristics to be analyzed, which have varied values (Kerlinger, 2006). Research variables are also the symbols to which we put values or scores (Kerlinger, 2006). Research variables include all aspects set by a researcher to be observed, about which information to be collected and put into conclusions (Sugiyono, 2009). This research included four variables as follows.

- (1) Service quality is influenced by five indicators, namely, reliability, responsiveness, empathy, assurance and tangibles (Parasuraman, 2001).

- (2) Consumer attachment is influenced by self Company connection, company behavior, customer behavior, brand characteristic, company characteristic and consumer brand characteristic (Yang and Kang, 2009).
- (3) Consumer satisfaction is influenced by the front office, infrastructure, guarantee and Corporate Social Responsibility (Ciptono, 2001).
- (4) Consumer loyalty is influenced by repurchasing, recommendation, endorse, negligence toward competitors and advocacy (Ciptono, 2001).

2.2.8 Research measuring scales. The pre-determined variables were measured using scales in the forms of questionnaires. These followings are commonly-used scales in research (Riduwan, 2009).

2.2.8.1 Likert scale. The Likert scale measures respondents' attitudes and perceptions about certain social phenomena. The social phenomena should have been specifically determined by the researcher, which is later referred to as a research variable.

The responses toward the research instrument in the form of a Likert scale ranges from strongly positive to strongly negative responses. The responses on the Likert scale are scored as follows:

strongly agree (5), agree (4), unsure (3), disagree (2), strongly disagree (1).

2.2.8.2 Guttman scale. This scale allows a researcher to obtain firm answers between yes or no, true or false, positive or negative, etc. The data obtained in this scale are in the form of interval data or a dichotomized ratio (two alternatives).

While the Likert scale contains five intervals, the Guttman scale only contains two intervals "agree or disagree." The Guttman scale is used when a researcher expects to obtain a firm answer about a certain problem.

2.2.8.3 Differential semantic scale. The differential scale measures ones' attitude yet it is not in the form of multiple choices or checklist. Instead, this scale is formed in a continuum line in which the strongly positive response is located on the right side of the line, while the strongly negative one is on the left side or vice versa.

This scale is used to collect interval data. It is often used to measure a certain attitude or characteristic of a research object.

2.2.8.4 Thurstone scale. Thurstone scale is designed by selecting certain points in the form of an interval scale. Every point has a key score that provides values with an equal gap. The Thurstone scale contains 40-50 statements that are relevant to a research variable to be measured. The data are then assessed by 20-40 experts who will judge the suitability of the statements with the content or construct to be measured.

2.2.8.5 Stapel scale. Stapel scale is the modification of the differential semantic scale. However, this scale provides no neutral options. It provides only positive or negative responses. This scale is a unipolar scale in which adjectives are only put in one pole.

The types of this scale remain debatable among experts. The data captured by this scale are interval data, yet as there are doubts regarding the gaps between items, the data are better regarded as ordinal data (Sekaran, 2000). The differential semantic scale was used in this research as this research compared two poles of different attitudes ("strongly satisfied" and "strongly dissatisfied"). Therefore, the results of the questionnaires were used to determine the level of consumer satisfaction among respondents. It was important to design a questionnaire that is easy to be understood and eases the respondents in giving their answers. Thus, a number of tests were conducted on the research instrument.

2.2.9 Validity test. A validity test was conducted to make sure that the research instrument measured the intended concept. The validity test was performed by correlating the score of each response to the total score (Sugiyono, 2009).

Corrected item-total correlation formula was used in this test as follows:

$$r_{i(x-i)} = \frac{r_{ix}S_x - S_i}{\sqrt{(S_x^2 + S_i^2 - 2r_{ix}S_iS_x)}} \quad (5)$$

where:

$r_{i(x-i)}$ = correlation coefficient of item i with a total score of all items (except item i);

$r_{i(x-i)}$ = correlation coefficient of item i with a total score;

S_i = standard deviation of item i ; and

S_x = standard deviation of the total score.

If the corrected item-total correlation value is found greater than 0.3, the instrument can be regarded valid (Solimun, 2010).

2.2.10 Reliability test. Reliability shows that an instrument can be trusted as a data collector for it has been considered good (Arikunto, 2006). Reliability relates to the degree of consistency of measurement results. A questionnaire is stated reliable when its results remain similar when a measurement is repeated to a similar object at different times. The reliability in the form of Cronbach's Alpha was tested in this research using this following formula:

$$r_{11} = \left[\frac{k}{(k-1)} \right] \left[1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right] \quad (6)$$

where:

k = number of items;

σ_b^2 = standard deviation of items; and

σ_t^2 = standard deviation of total.

If Cronbach's Alpha coefficient is found at $(r_{11}) \geq 0.7$, the instrument is regarded reliable.

2.2.11 Validity test. The materials and this research regarded primary data obtained from the questionnaires distributed to the costumers of PT Pelindo I, which included information about service quality, attachment, customer satisfaction and customer loyalty.

The population of this research was all customers of PT Pelindo I. The number of population in this research was infinite as the exact number of customers could not be determined. Samples were selected using non-probability sampling in the form of an accidental sampling method.

These followings are criteria of research respondents in this research:

- Respondents are passengers of PT Pelindo I.
- Minimum age of 17.
- Respondents are regular customers.

There were 107 samples selected as respondents of this research from 3 main ports of PT Pelindo I including Belawan Port, Dumai Port and Tanjung Pinang Port.

This research was conducted systematically based on the following steps:

- Determining research variables.
- Designing questionnaires based on the research variables determined.
- Determining the research population and samples.

- Administering pre-test to test the research instrument.
- Testing the instrument.
- Conducting evaluation to revise the questionnaire based on the results of the instrument testing.
- Collecting research data by distributing the questionnaire to the predetermined samples.
- Performing cluster analysis to the obtained data.

These followings are the steps used in performing the hybrid cluster analysis using R software:

- Performing average linkage clustering from the data.
- Measuring the Euclidian gap between the initial cluster.

Forming the clusters based on the closeness of Euclidian gaps between the initial cluster using the K-means hierarchy method as designed.

3. Methodology

3.1 Research data

Data used were primary ones obtained from the results of a questionnaire distributed to customers of PT Pelindo I reviewed from the variable of service quality, engagement, customer satisfaction and customer loyalty of PT Pelindo I.

3.2 Time and location

This research was conducted from September-November 2018. There were several activities done during these three months, namely, determining variables, arranging questionnaires, testing the research instruments, distributing questionnaires and analyzing the results of the questionnaires. The location was in three main ports of PT Pelindo I, namely, Belawan, Dumai and Tanjung Pinang.

3.3 Population and sample

In this study, the population was all service users of PT Pelindo I. This population is infinite because the service users of PT Pelindo I cannot be determined in exact numbers every day. The sampling of this study was non-probability with the accidental sampling method.

The respondent criteria are as follows:

- Passengers transport user of PT Pelindo I.
- Respondents are at least 17 years old.
- Passengers financing their own departure.

The sample of this study was 107 respondents from 3 major ports, namely, Belawan, Dumai and Tanjung Pinang.

4. Results and discussion

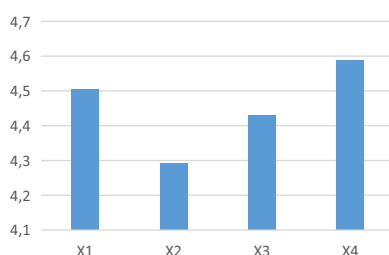
4.1 Descriptive analysis

A descriptive analysis was performed to obtain a general picture of the four research variables for each port and PT Pelindo, in general. The results of the descriptive analysis of the satisfaction variables are described in [Figure 1](#), that shows that the average score (mean)

of variable $X1$ and $X4$ is regarded very high (average score > 4.50). While, the average score of variable $X2$ and $X3$ is regarded high (average score between 3.51-4.50). Overall, the average score obtained for Belawan is 4.452, which is considered high (average score between 3.51-4.50).

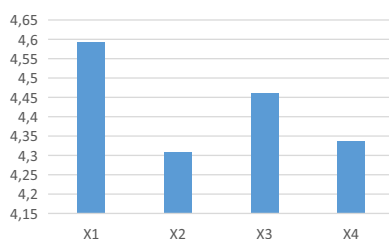
Figure 2 shows that the mean value of the $X1$ variable is regarded very high (average score > 4.50). While the scores of variables $X2$, $X3$ and $X4$ are categorized high (average score between 3.51-4.50). Overall, the average score of all variables for Dumai Port is categorized high at 4.424 (average score between 3.51-4.50).

Figure 3 shows that the mean scores of all variables are categorized high (between 3.51-4.50). The overall average score of Tanjung Pinang Port is 3.994, which is included in the high category (average score between 3.51-4.50).



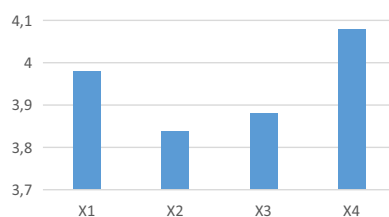
Source: Own elaboration

Figure 1.
The average score of
variable $X1$ - $X4$ in
Belawan port



Source: Own elaboration

Figure 2.
The average score of
variable $X1$ - $X4$ in
Dumai port



Source: Own elaboration

Figure 3.
The average score of
variable $X1$ - $X4$ in
Tanjung Pinang port

Finally, [Figure 4](#) shows the average score of all variables is categorized high (average between 3.51-4.50). Overall, the average score of PT Pelindo I is 4.270, which is included in the high category (average score between 3.51-4.50).

4.2 Cluster analysis

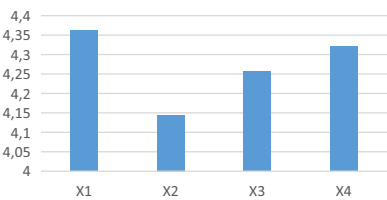
The purpose of this research was to segment the customer satisfaction of PT Pelindo I using the hybrid method. Primary data obtained from the customers of PT Pelindo I as respondents were regarded in this research. After carrying out the hybrid method segmentation process at the three main ports of PT Pelindo I and a combination of the three, optimal clusters were obtained as [Table 1](#) shows.

The results of segmentation showed that the optimal cluster in each research object was two clusters. Based on the interpretation of the five-point scale, a score greater than 4.5 is considered very high, while the one between 3.5-4.5 is within a high category. While a score between 2.5-3.5 is moderate and so on ([Solimun, 2010](#)). In this research, groups of customers with mean scores greater than 4.5 are considered a “special” customer. Customers with mean scores ranging between 3.5 and 4.5 are “good” customers and customers with mean scores ranging between 2.5 and 3.5 are “standard” customers. The results of the segmentation based on every port and the overall segmentation are explained in detail as follows.

4.3 Belawan port

The results of the segmentation done to Belawan Port are shown in [Figure 5](#). In the segmentation of the Belawan Port customers, the first two dimensions explain 91.4% of the total diversity and it seems that there have been two clearly visible. [Table 2](#) explains the score of the centroid variable of each cluster, and [Figure 6](#) illustrates the centroid scores.

Figure 4.
The average score of
variable *X1-X4* of PT
Pelindo I



Source: Own elaboration

Table 1.
Descriptive results of
segmentation per
port

Port	Cluster	Member
Belawan	1	16
	2	14
Dumai	1	38
	2	3
Tanjung Pinang	1	23
	2	13
Pelindo I	1	77
	2	30

Source: Own elaboration

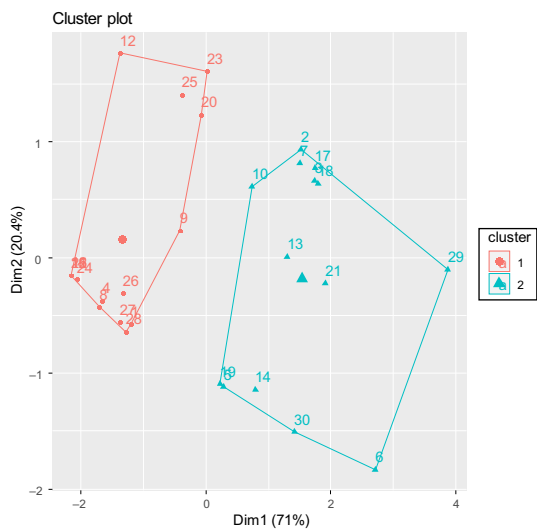


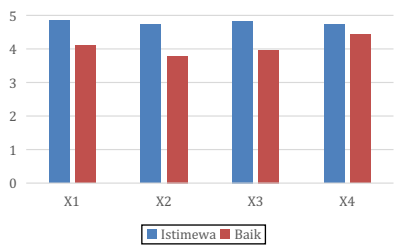
Figure 5.
Scatter plot cluster of
Belawan port

Source: Own elaboration

Variable	Cluster 1	Cluster 2
X1	4.85	4.10
X2	4.72	3.79
X3	4.83	3.97
X4	4.72	4.42

Source: Own elaboration

Table 2.
The centroid of
Belawan port



Source: Own elaboration

Figure 6.
The chart bar of
Belawan port
centroid cluster

Based on [Figure 6](#), it is understood that the highest service quality (*X1*), satisfaction (*X2*), attachment (*X3*) and loyalty (*X4*) are found in Cluster 1. Cluster 1 can be said as a “special” customer cluster with high variable scores (average score > 4.5). While the average service quality (*X1*), satisfaction (*X2*), attachment (*X3*) and loyalty (*X4*) in Cluster 2 are below the average perception of Cluster 1. Cluster 2 is considered a “good” customer cluster with high variable scores (average score between 3.5-4.5).

The “special” customer cluster consists of 16 respondents, while the “good” customer cluster consists of 14 respondents. Therefore, it can be concluded that 53% of Belawan Port customers are “special” customers and the remaining 47% are “good” customers.

4.4 Dumai port

Figure 7 shows the centroid score of customer satisfaction in Dumai Port for every cluster. In the customer segment of Dumai Port, the first two dimensions explain 70.8% of the total diversity and it appears that there have been two clearly visible segmentations.

The centroid score of the variable in each cluster is explained in Table 3, and the scores are shown in Figure 8.

Figure 8 shows that the highest service quality (X1), satisfaction (X2), attachment (X3) and loyalty (X4) are found in Cluster 1. Cluster 1 can be regarded a “special” customer cluster even the scores are not very high (average score > 4.5). The scores of the Cluster 1 is higher than Cluster 2, making it considered as a “special” cluster. Meanwhile, the average service quality (X1), satisfaction (X2), attachment (X3) and loyalty (X4) in Cluster 2 are below the average perception of Cluster 1. Cluster 2 is a “good” customer cluster with a high variable score (average between 3.5-4.5). The “special” customer cluster consists of 38 respondents, while the “good” customer cluster consists of 3 respondents. Hence, 93% of Dumai Port customers are “special” customers and the remaining 7% are “good” customers.

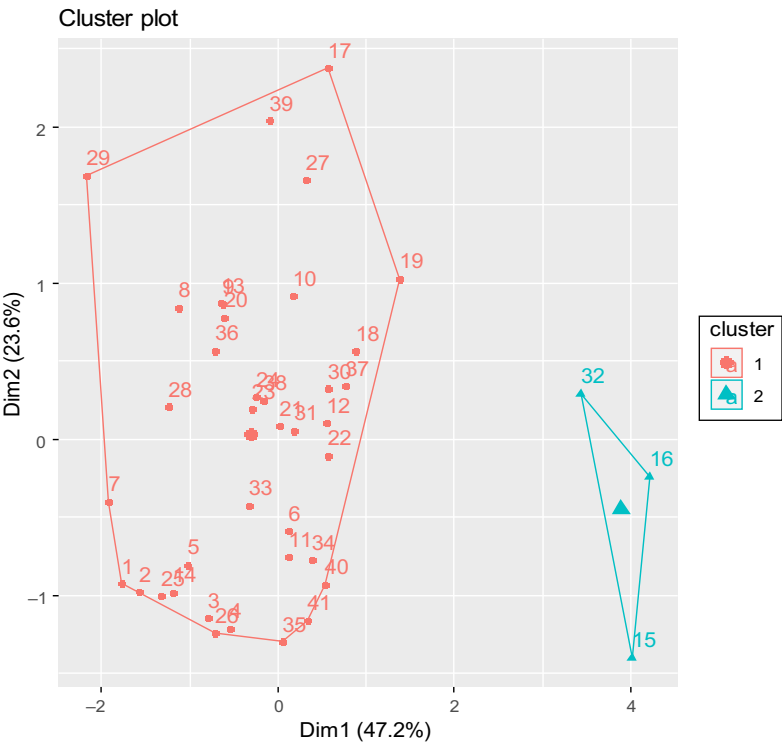


Figure 7.
Scatter plot cluster of
Dumai port

Source: Own elaboration

4.5 Tanjung pinang port

Figure 9 illustrates a centroid graph of customer satisfaction in Tanjung Pinang Port for each cluster.

In the customer segment of the Tanjung Pinang Port, the first two dimensions explain 89.8% of the total diversity and it indicates two clearly visible segmentations. The variable centroid scores of each cluster are explained in Table 4.

Figure 10 explains those centroid scores.

Based on Figure 10, it can be seen that the highest average service quality (X1), satisfaction (X2), attachment (X3) and loyalty (X4) is found in Cluster 1. Cluster 1 is regarded as a “good” customer cluster with high variable scores (average score between 3.5-4.5). While the average service quality (X1), satisfaction (X2), attachment (X3) and loyalty (X4) in Cluster 2 are below the average perception of Cluster 1. Cluster 2 is a “standard” customer cluster with a variable moderate score (average between 2.5-3.5). The “good” customer cluster consists of 23 respondents, while the “standard” customer cluster consists of 13 respondents. It indicates that 64% of Tanjung Pinang Port customers are “good” customers and the remaining 36% are “standard” customers.

4.6 Pt pelindo I

Figure 11 describes the centroid score of the customer satisfaction of PT Pelindo I for each cluster.

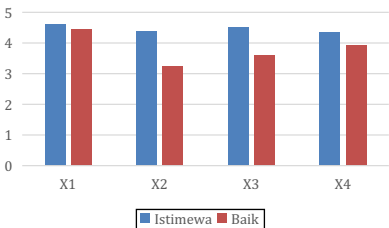
In the segmentation of PT Pelindo I customers, the first two dimensions explain 88.9% of the total diversity and there have been two clearly visible segments. Table 5 explains the variable centroid scores of each cluster, which are shown in Figure 12.

Based on the graph above, the highest average score of service quality (X1), satisfaction (X2), attachment (X3) and loyalty (X4) is found in Cluster 1. Cluster 1 can be regarded as a “special” customer cluster with very high variable scores (>4.5). While the average service quality (X1), satisfaction (X2), attachment (X3) and loyalty (X4) in Cluster 2 are below the

Variable	Cluster 1	Cluster 2
X1	4.60	4.44
X2	4.39	3.26
X3	4.52	3.61
X4	4.36	3.93

Source: Own elaboration

Table 3.
The centroid of
Dumai port



Source: Own elaboration

Figure 8.
The chart bar of
Dumai port centroid
cluster

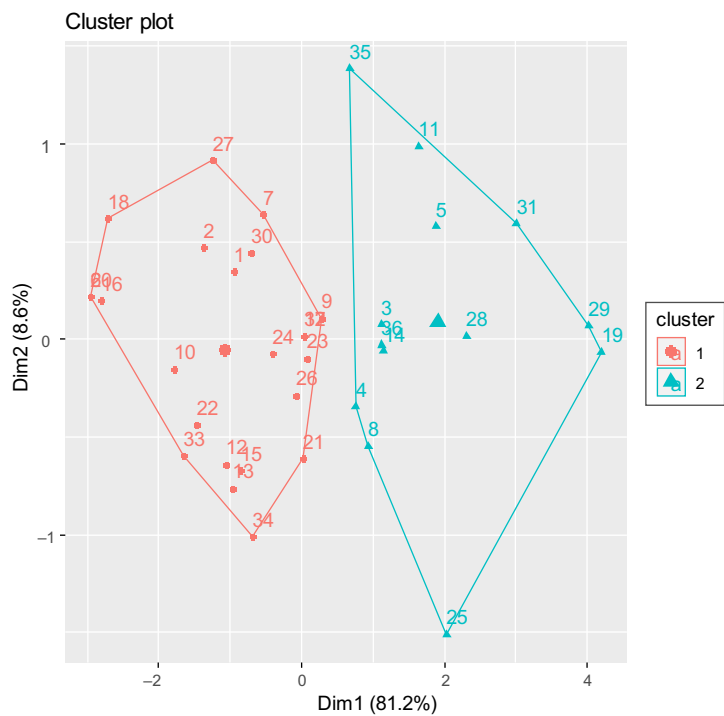


Figure 9.
Scatter plot cluster of
Tanjung Pinang port

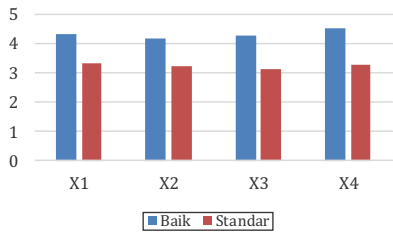
Source: Own elaboration

Table 4.
The centroid of
Tanjung Pinang port

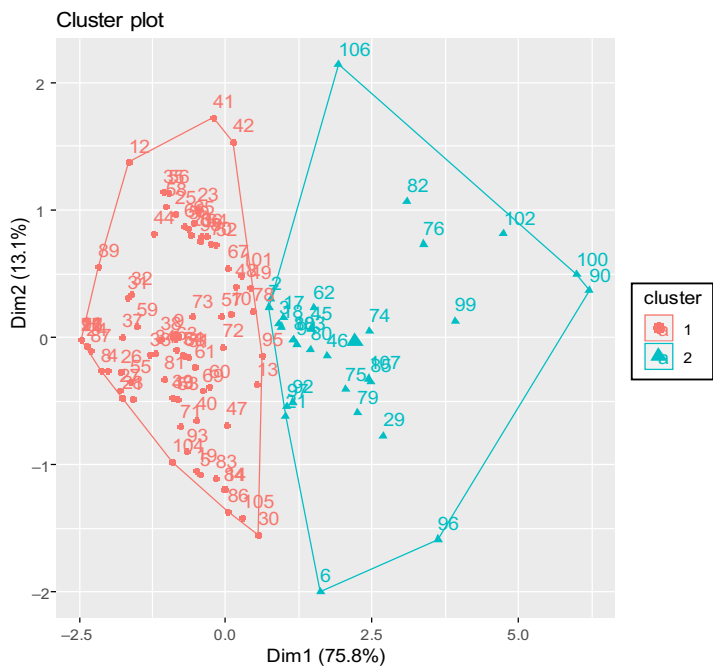
Variable	Cluster 1	Cluster 2
$X1$	4.34	3.32
$X2$	4.17	3.23
$X3$	4.28	3.16
$X4$	4.53	3.27

Source: Own elaboration

Figure 10.
The chart car of
Tanjung Pinang port
centroid cluster



Source: Own elaboration



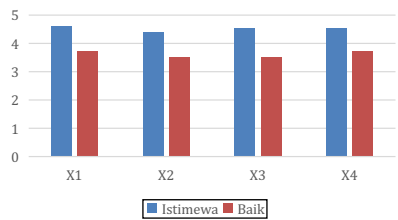
Source: Own elaboration

Figure 11.
Scatter plot cluster of
PT Pelindo I

Variable	Cluster 1	Cluster 2
X1	4.60	3.72
X2	4.39	3.50
X3	4.54	3.50
X4	4.54	3.74

Source: Own elaboration

Table 5.
The centroid of PT
pelindo I



Source: Own elaboration

Figure 12.
The chart bar of PT
Pelindo I centroid
cluster

average perception of Cluster 1. Cluster 2 is a “good” customer cluster with a high variable value (average between 3.5-4.5).

The “special” customer cluster consists of 77 respondents, while the “good” customer cluster consists of 30 respondents. Therefore, 72% of PT Pelindo I customers are “special” customers and the remaining 28% are “good” customers.

5. Conclusions

Based on the results of this research, conclusions were drawn as follows:

- The cluster hybrid analysis performed in this research resulted in two segmentations in Belawan Port, Dumai Port, Tanjung Pinang Port and in PT PELINDO I, in general.
- The customer segmentation of Belawan Port shows that 53% of customers perceived the quality service for passengers offered by the port special, while the rest 47% find the service good.
- The customer segmentation of Dumai Port shows 93% of customers find the service offered by Dumai Port special, while 7% of them find the service good.
- The customer segmentation of Tanjung Pinang Port shows 64% of customers find the service offered by Tanjung Pinang Port special, while 36% of them find the service standard.
- The overall customer segmentation of PT Pelindo I shows that 72% of customers find the passenger service offered by PT Pelindo special, while 28% of them find the service good.

Acknowledgement

The authors contributed to the paper in the following ways:

- Cahyana, Bambang Eka
Corresponding Author
Conceptualization (Equal)
Data curation (Equal)
Formal analysis (Equal)
Funding acquisition (Equal)
Investigation (Equal)
Methodology (Equal)
Project administration (Equal)
Supervision (Equal)
Validation (Equal)
Visualization (Equal)
Writing-original draft (Equal)
Writing-review & editing (Equal)
- Nimran, Umar
Conceptualization (Equal)
Data curation (Equal)
Project administration (Equal)
- Nayati, Hamidah
Formal analysis (Equal)
Methodology (Equal)
- Iqbal, Mohammad

References

- Arikunto, S. (2006), *Prosedur Penelitian Suatu Pendekatan Praktek*, PT. Rineka Cipta, Jakarta.
- Cheu, E., Keongg, C. and Zhou, Z. (2004), "On the two-level hybrid clustering algorithm", *International Conference on Artificial Intelligence in Science and Technology*.
- Chipman, H. and Tibshirani, R. (2005), "Hybrid hierarchical clustering with applications to microarray data", *Biostatistics*, Vol. 7 No. 2, pp. 286-301.
- Ciptono, F. (2001), *Strategi Pemasaran*, Andi Offset, Yogyakarta.
- Hair, F., Black, W., Babin, B. and Anderson, R. (2010), *Multivariate Data Analysis*, 7th ed., Pearson, New York, NY.
- Kerlinger (2006), *Asas-Asas Penelitian Behavioral*, Gadjah Mada University Press, Yogyakarta.
- Kotler and Keller (2013), "Manajemen pemasaran", *Jilid I Edisi Ke, 13*, Erlangga, Jakarta.
- Latan, H. (2014), *Aplikasi Analisis Data Statistik Untuk Ilmu Sosial Sains Dengan Stata*, Bandung: Alfabeta.
- Parasuraman (2001), "The behaviorial consequences of service quality", *Journal of Marketing*, Vol. 60 No. 2, p. 60.
- Riduwan (2009), *Skala Pengukuran Variabel-Variabel Penelitian*, Alfabeta, Bandung.
- Sekaran, U. (2000), *Research Methods for Business: A Skill Building Approach*, John Wiley and Sons, Singapore.
- Solimun (2010), *Analisis Multivariat Pemodelan Struktural Metode Partial Least Square-PLS*, CV. Citra, Malang.
- Sugiyono (2009), "Metode penelitian kuantitatif", *Kualitatif Dan R&D*, Alfabeta, Bandung.
- Sukbukti, R. (2017), *K-Means Clustering Dan Average Linkage Dalam Pembentukan Portofolio Saham*, Seminar Matematika UNY.
- Yang, S. and Kang, M. (2009), "Measuring blog engagement: testing a four-dimensional scale", *Public Relations Review*, Vol. 35 No. 3.

Corresponding author

Bambang Eka Cahyana can be contacted at: bambang_eka@student.ub.ac.id