Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

Analysis of Passenger Satisfaction Levels Using the K-Means Cluster and Hierarchial Cluster Methods in Purabaya Sidoarjo Type A Terminal Services

Zahra Khania Putri[1*], Akmal Suryadi[2]

[1,2]Fakultas Teknik & Sains, Universitas Pembangunan Nasional "Veteran" Jawa Timur
22032010106@student.upnjatim.ac.id[1], akmal.suryadi65@gmail.com[2]*

Abstract

Purabaya Type A Terminal is one of the terminals that provides public transportation facilities in the form of bus transportation ranging from city buses, intra-provincial intercity buses, and inter-city inter-provincial buses. This study aims to analyze the level of passenger satisfaction with the service at the terminal and group each variable into a homogeneous group. The research methods used are the K-means Cluster and Hierarchial Cluster methods. The data from the questionnaire will be grouped into several groups that have relatively homogeneous properties using the help of SPSS software. From the results of the study, the output of Non-Hierarchial clusters was obtained in the form of Initial Cluster Centers, Iteration History, Final Cluster Centers, ANOVA, and Number of Cased In Each Cluster. Meanwhile, the output of the Hierarchial cluster is in the form of Case Processing Summary and Dendogram Using Average Linkage. Through analysis using the cluster method, all variables were obtained including in cluster 1 and none were included in cluster 2, the iterations carried out were 2 times and the valid data value was 100 data with a missing value of 0.

Keywords: *Hierarchial Cluster, K-Means Cluster, Purabaya Type A Terminal, SPSS*

1. Introduction

Statistics is a mathematical science that is widely used in daily life. Statistics is a science related to the method of data collection, processing or analysis as well as drawing conclusions based on data sets and analyses carried out [1]. In statistics, there is a branch that focuses on data analysis. One of the methods that is often used in the industrial world is cluster analysis. This technique aims to group objects into specific categories based on the similarities or homogeneous characteristics they have. To simplify the analysis process in a case study, software such as SPSS is often used. SPSS is one of the software designed to support statistical analysis, allowing data processing more efficiently and producing accurate output.

Purabaya Type A Terminal is one of the type A terminals in Indonesia, precisely in Sidoarjo Regency, East Java Province. This terminal provides public transportation facilities in the form of bus transportation ranging from city buses, intra-provincial intercity buses, and inter-provincial intercity buses. The terminal is one of the places that is quite in demand by passengers because many passengers are interested in using public transportation buses. Generally, before leaving for the destination city, passengers stop for some time at the terminal to wait for the bus they want to use. Improvements in services carried out by officers need to be done to provide comfort to passengers. Thus, it is necessary to make observations on the assessment of the level of passenger satisfaction with service through a questionnaire to passengers of the Purabaya Type A Terminal which will then be analyzed data. This study aims to analyze the level of passenger satisfaction with the service at the Type A Purabaya Terminal

To obtain data to be analyzed, a questionnaire was distributed to passengers of the Purabaya Type A Terminal. The method used to solve the problem in this study is cluster analysis which includes the k-means cluster test and the cluster hierarchial test. The k-means cluster test is a cluster analysis test that groups objects by being able to determine their clusters first. Meanwhile, the cluster hierarchial analysis test is an analysis test that groups objects gradually and stratively based on their similarity, but the number of clusters is not yet known. To facilitate data processing and analysis, SPSS (Statistical Product and Service Solutions) software is used. SPSS is a software that functions to perform advanced statistical analysis, data processing using machine learning algorithms, string analysis, and data analysis that can be integrated with other data analysis platforms. SPSS is one of the computer programs specifically designed to process data using certain statistical methods [2].

Purabaya Type A Terminal is one of the terminals that provides public transportation facilities in the form of bus transportation ranging from city buses, intra-provincial intercity buses, and inter-city inter-provincial buses. Improvements in services carried out by officers need to be done to provide comfort to passengers. This study aims to determine the level of passenger satisfaction with the service of the Purabaya Type A Terminal by examining the quality of passenger service through the observation of questionnaires distributed to passengers. This study uses cluster analysis, where this analysis includes cluster k-means and hierarchial cluster tests to analyze data. SPSS (Statistical Product and Service Solutions) software is used to assist in the analysis of this data. So that the data obtained can be processed faster and get accurate results.

2. Literature Review

2.1. SPSS Software

SPSS (Statistical Package for the Social Sciences) is a very powerful Windows-based statistical analysis software. The menu interface and dialogs in SPSS are designed to provide an intuitive user experience. This software is very useful in various fields of research. As one of the most popular statistical data processing software in the world, SPSS is widely used in market research, quality control, and scientific research. With a graphics-based data management system, SPSS provides descriptive menus and simple dialogues, making it easy for users to operate and understand. SPSS is one of the most in-demand applications by analysts and researchers to process statistical data [3].

2.2. Multivariate Data Analysis

Multivariate analysis is a statistical technique that is useful for analyzing many variables simultaneously. Multivariate analysis is divided into two types, namely the dependency and independence methods. It relates to the data being analyzed and the purpose of the analysis. Some of the methods used in multivariate statistical analysis include main component analysis, factor analysis, canonical correlation analysis, multivariate regression analysis, multivariance analysis, discrimination analysis, and cluster analysis [4].

2.3. Cluster Analysis

Cluster analysis is a method used to form subgroups of individuals or objects that have similarities. The main goal is to group a sample of entities into several small groups that are mutually exclusive based on similarities between them. Unlike other methods, cluster analysis does not have predetermined groups, but rather aims to identify those groups.

The cluster analysis process consists of three main stages. First, it measures the degree of similarity or relationship between entities to determine the number of groups present in the sample. Second, carry out the grouping process by dividing entities into several groups or clusters. Third, compile a profile for each group or variable to understand the composition of each. This profile can often be created using discrimination analysis of groups that have been generated from clustering techniques [5].

There are two basic assumptions that must be met, namely the adequacy of the sample and the absence of multicollinearity among the variables. The assumption of sample adequacy is met if the Kaiser-Meyer-Olkin (KMO) value is greater than 0.5.

$$\text{KMO} = \frac{\sum_{i \neq j} \sum r_{ij}^2}{\sum_{i \neq j} \sum r_{ij}^2 + \sum_{i \neq j} \sum a_{ij}^2} \quad (1)$$

Description:
r_{ij}^2 : Correlation between variables i and j
a_{ij}^2: Partial correlation between variables i and j

Checking the assumption of nonmulticollinearity is carried out by the Bartlett test with the initial hypothesis, test statistics, and test criteria as follows:

Hypothesis:
H0 : R = 1
H1 : R ≠ 1

Test statistics:

$$\text{Barlettt} = \ln |R| \left(n - 1 - \frac{2p+5}{6}\right) \quad (2)$$

|R| : Correlation matrix determinant values
n : Many observations
p : Many variables

2.4. Hierarchial Cluster

The hierarchy method is a cluster analysis technique that is carried out in stages and tiers, resembling a tree structure. This method produces a series of partitions by combining or separating clusters optimally at each stage, based on certain criteria. The results of the analysis using this method are usually displayed in the form of a dendogram, which is a visual representation that shows the cluster formation process as a whole. In addition, each stage includes a distance coefficient value calculated based on a data pair. These distance measurements can be done with a variety of approaches, such as the Euclidean distance or the Manhattan distance. Hierarchical methods are divided into two

types, namely the agglomerative method (from bottom to top) and the divisive method (from top to bottom), which differ in the direction of the grouping process. There are four agglomerative methods in cluster formation, namely [6]:

1. Single linkage, which is the distance between two clusters measured by the closest distance between an object in one cluster and an object in another cluster.
2. Complete linkage, which is the distance between two clusters measured by the farthest distance between an object in one cluster and an object in another cluster.
3. Average linkage, which is the distance between two clusters measured by the average distance between an object in one cluster and an object in another cluster.
4. Word's method, which is an agglomerative method that is often used in continuous data is the Ward method. This approach combines clusters if the total inequality squared with the center of the cluster is minimum across all possible merge options.

2.5. K-Means Cluster

K-means is a method that divides data into several groups with the aim that data that have the same characteristics are in the same cluster and vice versa, data that has different characteristics are in different clusters [7]. K-means cluster has a directionless nature because the analysis in the grouping does not have a definite cluster. There are two fundamental analyses in finding a solution from the k-means cluster. First, the solution must come from many predefined clusters. Second, the selection of centroid clusters is not very reliable, so clustering depends on the centroid chosen. There are many programs that start with the selection of the first case k (k = number of clusters) as the initial cluster center. Thus, the cluster results depend on data observation [8].

3. Research methodology

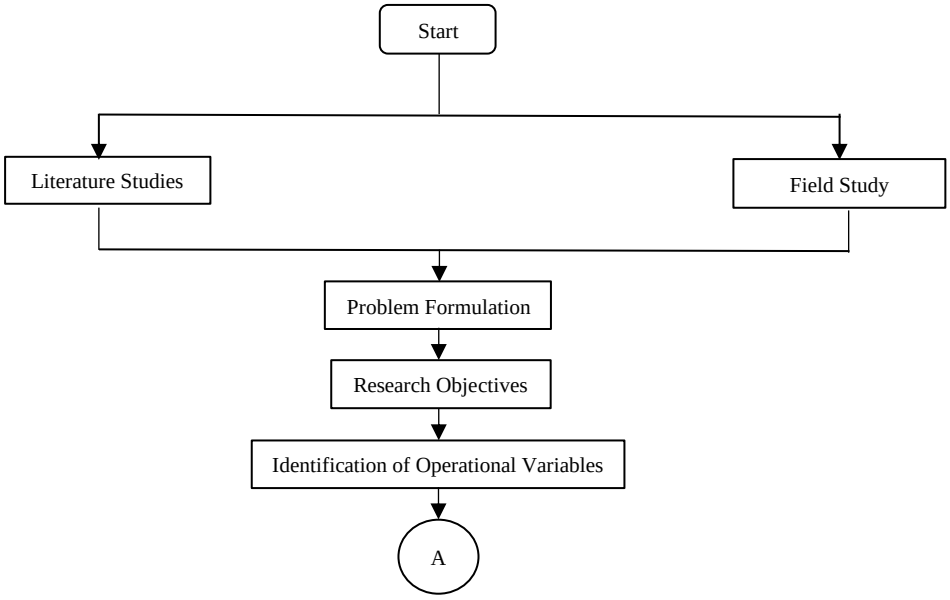
3.1. Data collection stage

Data collection is an activity to collect information from a source that is relevant to a research. The data that has been collected is used as input at the data processing stage. The data collection method used in this study is a quantitative method. The data needed, namely:

1. Respondent name data
2. Adequate terminal facility assessment data
3. The assessment data of the information provided by the terminal is appropriate
4. Assessment data on the number of officers that are adequate
5. Arrival and departure schedule assessment data is easily accessible
6. Guaranteed terminal security assessment data
7. Assessment data on the absence of discrimination
8. Assessment data on the absence of officer fraud
9. Assessment data no gratuities for terminal officers
10. Assessment data there is no illegal levy
11. Assessment data without any transactions

3.2. Data processing stage

Data processing is a process of collecting, processing, and analyzing data to produce useful information. The steps to solve the problem in this study are as follows:



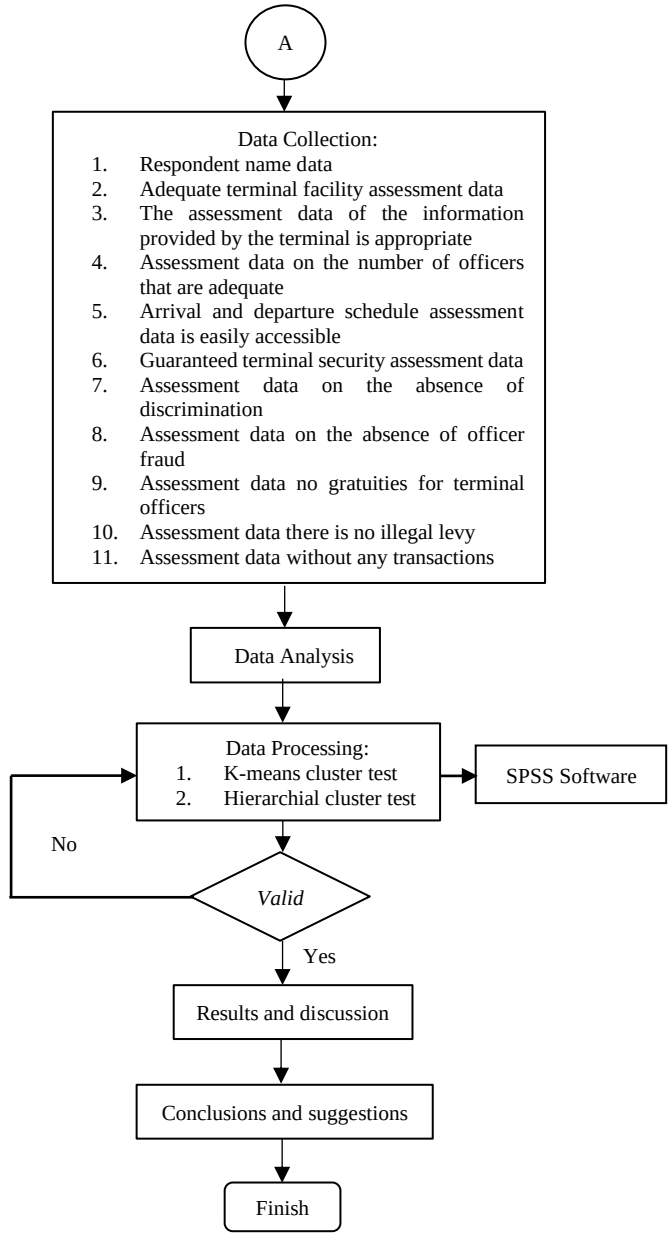


Fig. 1: Flowchart

Based on the flowchart above, the following are the troubleshooting steps:

1. Start
   Flowchart begins to be interpreted as a sign that the research activity has started and the next steps are taken. The flowchart began to be symbolized by a terminator that showed the beginning of the research process. The small arrows of the flowchart point to the steps that must be taken to start the research process.
2. Literature Studies
   The process of collecting literature reviews, reading and taking notes, and processing research materials. This literature study aims to find information and examine basic theories related to the problem being researched. The sources used are by using books and journals.
3. Field Study
   The process of collecting data and determining case studies or problems to be researched. This activity can be in the form of observation from the literature review obtained or it can also be in the form of direct observation in the field.
4. Problem Formulation
   Problem formulation is the formulation or presentation of the problem to be researched in a study or study that can help determine the direction of the research and help clarify the goals and objectives of the research being conducted. The formulation of the problem for this study is related to the analysis of the K-means cluster test and the Hierarchial test of the passenger satisfaction level of the Purabaya Sidoarjo Type A Terminal service using SPSS software.
5. Research Objectives
   The purpose of the research is to identify the concept of why the research is carried out and to explain or predict the solution to the problem being researched. The purpose of the study is to group variables that have relatively similar properties into a homogeneous group.

6. Identification of Operational Variables
   This step aims to identify what variables are used and how they function. There are two types of variables, independent variables and bound variables. The independent variables used in this study are passenger assessment data of Purabaya Type A Terminal which includes adequate terminal facilities, appropriate information provided, adequate number of officers, easily accessible schedules, guaranteed terminal security, no discrimination, no officer fraud, no gratuities, and no brokerage. Meanwhile, the bound variable is an assessment of the level of passenger satisfaction with the service at the Purabaya Type A Terminal.
7. Data Collection
   Data collection is an activity to find data in the field that will be used to answer research problems. The data collected for this study was obtained from questionnaires distributed to passengers. The data collected, namely the assessment data of terminal facilities that are adequate, the information provided is appropriate, the number of officers is adequate, the schedule is easily accessible, the security of the terminal is guaranteed, there is no discrimination, there is no cheating of officers, there is no gratuity, there is no illegal levy, and there is no scam.
8. Data Analysis
   Data analysis is a technique that is carried out in a process of processing data that has been processed. The data analysis used in this study was by using cluster analysis, k-means cluster test and cluster hierarchial test.
9. Data Processing
   Data processing is the process of processing raw data into information that is easily accepted. In this step, SPSS software is used by using k-means cluster test analysis and cluster hierarchial test and by using a data collection method with a quantitative method.
10. Valid
    Valid can be interpreted as data that has been processed in accordance with procedures and has shown results. Since the data processing process shows valid results, the flowchart arrows point to the next step. If the data obtained is invalid, then the arrow will Return to data processing.
11. Results and Discussion
    The results of this study for the k-means cluster test were obtained initial cluster center, iteration history, final cluster center, ANOVA table, and number of cases in each cluster. Meanwhile, for the hierarchial cluster test, case processing summary and dendogram were obtained.
12. Conclusions and Suggestions
    Conclusion means concluding all research results starting from the methods and software used and the results of data processing that show valid and reliable data or not. Suggestions on research are also carried out to improve and develop research results to be better.
13. Finish
    The completed flowchart shows that all research activities have been completed and the desired output has been obtained.

4. Results and Discussion

4.1. Data processing

a. Raw data

Purabaya Type A Terminal is one of the terminals that provides public transportation facilities in the form of bus transportation ranging from city buses, intra-provincial intercity buses, and inter-city inter-provincial buses. Improvements in services carried out by officers need to be done to provide comfort to passengers. For this reason, the terminal wants to know how the level of passenger satisfaction with the service at the Purabaya Type A Terminal is. A list of questionnaires was distributed to passengers who contained questions about their opinions and assessments of the variables about services at the Purabaya Type A Terminal with respondents' answers in the form of assessments with measurements in the form of a Likert scale. The following is a recap of the questionnaire results:

Table 1: Questionnaire result data recap

Name	Terminal facilities	Information	Number of officers	Bus Schedule	Security	Discrimination	Officer fraud	Gratuities	Illegal levies	Brokerage
Tri Joko	5	5	5	5	4	4	4	4	5	4
Erna	2	2	2	3	2	2	3	2	2	2
Raisya	5	5	5	5	5	5	5	5	5	5
Najwa	5	5	5	5	4	2	2	5	4	4
Marchel	4	5	4	3	4	4	4	4	4	4
Dewi	4	5	3	4	5	5	4	5	5	4
Rani	5	4	4	5	3	3	4	4	5	4
Adit	4	5	4	5	4	5	4	4	3	3
Rizky	5	5	4	5	4	5	4	5	3	4
Milan	5	5	4	3	4	3	4	5	3	4
Fajar	4	3	4	3	4	4	5	4	3	4
Lina	5	4	5	4	5	4	3	4	3	5
Shezan	5	4	5	5	4	3	4	5	5	4
Syakira	5	4	4	5	5	4	3	4	3	4
Muthia	5	4	3	4	4	4	3	4	2	4
Delanda	4	5	4	4	3	4	5	4	3	4
Yustika	4	5	3	4	3	4	5	3	4	5
Dewi Atika	5	4	4	3	5	4	5	4	5	3
Nadhir Eka	4	5	3	4	5	3	4	5	4	3
Vina	5	4	5	3	4	4	5	4	3	4
Nurul	5	4	3	4	4	5	4	3	4	5
Uwais	5	4	3	4	4	4	4	3	4	3
Ihsan Arby	5	4	4	5	3	4	5	4	4	3
Chandra	3	4	4	5	5	5	4	4	4	3

Name	Terminal facilities	Information	Number of officers	Bus Schedule	Security	Discrimination	Officer fraud	Gratuities	Illegal levies	Brokerage
Iwan Fuad	4	5	4	5	3	4	5	5	4	3
Arga	3	5	4	4	5	3	4	3	4	5
Nadira	5	4	5	4	5	4	5	3	3	4
Damar Adit	4	5	5	4	5	4	4	3	3	4
Adinda	4	3	3	5	4	4	5	3	3	4
Haris	4	5	5	3	4	4	3	3	5	4
Belinda	3	5	4	5	5	4	3	5	3	3
Aliyyah	4	5	3	4	4	4	5	4	3	4
Iqbal	4	5	5	3	4	3	5	4	4	3
Arum	4	5	3	3	5	5	4	4	5	3
Siska	5	4	4	5	3	4	3	4	4	5
Taufik	4	3	4	5	4	4	3	5	4	4
Malik Kris	5	3	4	4	5	3	3	4	5	4
Bagus	5	4	3	4	3	5	4	5	3	3
Citra Okta	3	4	4	5	4	5	3	4	4	5
Bayu Adji	4	5	4	5	3	3	4	4	5	4
Mayang	3	5	5	4	3	4	4	5	3	5
Wulan Sari	5	3	5	3	4	4	4	3	5	5
Raina mela	4	5	4	3	4	4	5	4	3	5
Reyhan	3	5	4	3	5	5	4	5	3	4
Aminah	4	3	3	5	5	5	3	4	4	5
Kalya azza	5	4	5	4	4	3	4	3	4	5
Rizky Anan	4	4	4	4	5	4	3	3	3	4
Keanu	5	4	5	4	5	4	4	4	3	4
Fabian	4	5	4	3	4	4	5	3	3	4
Daffa	3	5	4	5	4	4	5	4	3	5
Kaira Alan	3	4	4	3	5	5	4	4	5	3
Ayla	4	5	5	4	3	4	3	4	4	5
Dimas	5	5	4	5	4	3	4	4	5	5
Shaqueena	2	4	4	5	3	5	4	4	5	3
Hendra	5	4	3	4	5	4	4	3	3	4
Mahardhika	5	4	4	4	3	4	4	5	4	3
Ibrahim	5	5	4	3	4	4	4	4	3	5
Eko	4	5	5	5	4	4	3	3	4	5
Nabila	5	4	4	5	3	5	4	4	5	3
Firdaus	4	4	5	4	4	5	4	3	4	5
Vian	4	5	5	4	4	5	4	4	4	5
Guna	4	5	4	4	4	5	5	4	5	4
Adnan Yuli	4	4	4	5	4	5	4	5	4	5
Leo Bima	3	5	4	5	5	4	4	5	3	4
Dhea	4	4	5	3	5	4	4	4	5	4
Daniel	5	3	4	3	5	5	4	4	4	4
Nazwa Siti	4	5	5	4	4	3	4	4	5	4
Damar Gali	4	5	4	4	4	4	3	4	4	5
Satrio	4	5	4	4	4	4	4	5	4	3
Sony	5	4	5	5	4	5	4	5	5	5
Alya	5	4	5	4	5	5	5	4	5	4
Rian	5	3	4	5	4	5	4	4	5	5
Maura diva	4	4	4	5	4	5	4	4	5	4
Della	4	5	5	4	4	5	4	4	5	5
Ferisha	4	5	3	4	3	5	4	4	5	5
Andhara	4	5	5	5	4	3	5	4	4	5
Cindy	4	5	5	4	4	5	4	5	4	4
Chandra	5	5	4	5	4	4	5	3	5	5
Java	4	5	5	4	5	4	5	4	5	5
Jendra	5	4	5	4	4	5	4	4	5	5
Jihan	4	5	4	5	4	5	4	5	5	4
Nagisa	4	5	4	4	5	5	4	5	5	3
Gema	5	5	3	5	4	4	4	5	3	5
Ghina	4	5	5	4	3	4	5	3	5	4
Ivana Amel	5	3	5	5	5	4	4	4	5	3
Fathan	5	4	4	5	5	3	5	4	4	4
Haris	5	5	4	4	3	4	4	4	5	5
Anna	5	4	5	5	5	4	3	4	4	5
Siti Zahra	5	4	5	4	5	5	5	5	4	5
Jeana	4	5	5	4	5	4	5	5	3	5
Tiara	4	5	4	5	5	5	4	4	5	3
Rendi Wina	4	5	4	5	5	5	5	4	4	3
Sabrina	5	5	4	4	4	4	5	4	5	5
Diana Ayu	5	5	3	3	4	4	5	5	3	3
Iliana	4	5	5	3	4	5	4	5	5	4
Hanan Juni	4	5	4	5	4	5	4	4	3	4
Uwi	5	4	5	4	4	3	4	5	2	4
Azka	4	5	3	4	5	4	5	4	4	5
Ragil Rena	5	4	4	5	5	4	3	3	2	3

Name	Terminal facilities	Information	Number of officers	Bus Schedule	Security	Discrimination	Officer fraud	Gratuities	Illegal levies	Brokerage
Laika	5	4	4	5	5	4	4	3	5	4

b. Input Data

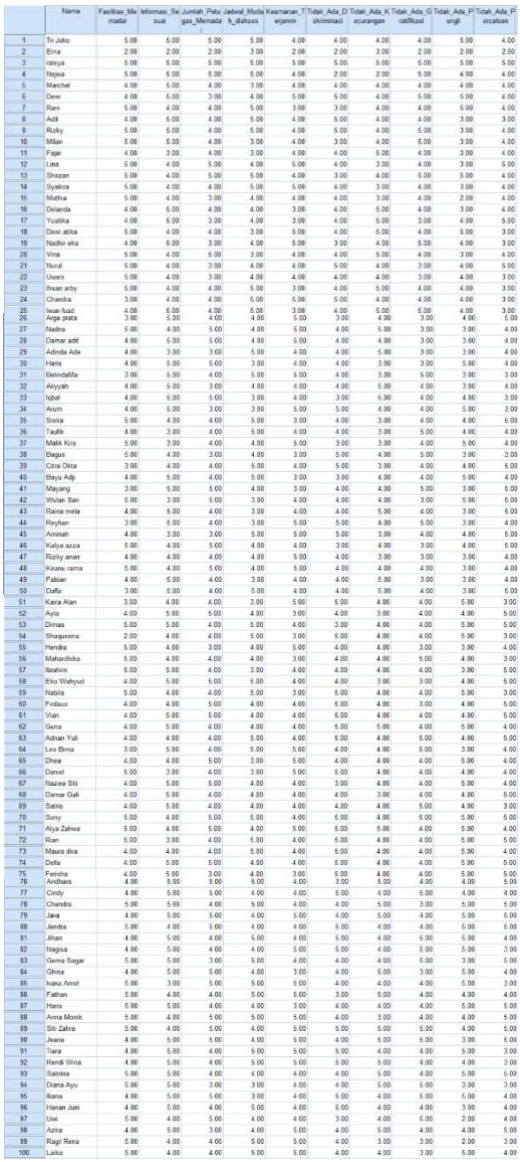


Fig. 2: Input Data on Passenger Satisfaction Level Questionnaire Results for Services at Terminal A Purabaya using SPSS Software

c. Output Data

1) Non-hierarchial cluster

Initial Cluster Centers

	Cluster	
	1	2
Fasilitas_Memadai	5.00	2.00
Informasi_Sesuai	5.00	2.00
Jumlah_Petugas_Memadai	5.00	2.00
Jadwal_Mudah_diakses	5.00	3.00
Keamanan_Terjamin	5.00	2.00
Tidak_Ada_Diskriminasi	5.00	2.00
Tidak_Ada_Kecurangan	5.00	3.00
Tidak_Ada_Gratifikasi	5.00	2.00
Tidak_Ada_Pungli	5.00	2.00
Tidak_Ada_Percaloan	5.00	2.00

Fig. 3: Initial Cluster Centers

Analysis:

The figure above is the output of the Initial Cluster Center which shows the results of the temporary process of data grouping carried out. Based on the output above, it can be seen that the clustering process is divided into 2 clusters.

Iteration History[a]

Iteration	Change in Cluster Centers	
	1	2
1	2.586	.000
2	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 2. The minimum distance between initial centers is 8.944.

Fig. 4: Iteration History

Analysis:
The output above is the Iteration History. The iterations carried out are as many as 2 iterations to achieve the optimum value. The shortest distance between cluster centers, which is 8,944

Final Cluster Centers

	Cluster	
	1	2
Fasilitas_Memadai	4.33	2.00
Informasi_Sesuai	4.45	2.00
Jumlah_Petugas_Memadai	4.18	2.00
Jadwal_Mudah_diakses	4.22	3.00
Keamanan_Terjamin	4.19	2.00
Tidak_Ada_Diskriminasi	4.18	2.00
Tidak_Ada_Kecurangan	4.10	3.00
Tidak_Ada_Gratifikasi	4.08	2.00
Tidak_Ada_Pungli	4.02	2.00
Tidak_Ada_Percaloan	4.14	2.00

Fig. 5: Final Cluster Centers

Analysis:
The output above is the Final Cluster Centers. The number of clusters formed is 2 clusters. It can be seen that all variables are included in cluster 1, namely the variables of adequate facilities, appropriate information, adequate number of officers, easily accessible schedules, guaranteed security, no discrimination, no fraud, no gratuities, no illegal levies, and no fraud. This is because all of these variables have higher values in cluster 1. Meanwhile, no variables go into cluster 2 because all variables have lower values in cluster 1.

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Fasilitas_Memadai	5.390	1	.469	98	11.483	.001
Informasi_Sesuai	5.965	1	.434	98	13.739	.000
Jumlah_Petugas_Memadai	4.713	1	.477	98	9.884	.002
Jadwal_Mudah_diakses	1.479	1	.542	98	2.729	.102
Keamanan_Terjamin	4.756	1	.483	98	9.844	.002
Tidak_Ada_Diskriminasi	4.713	1	.497	98	9.478	.003
Tidak_Ada_Kecurangan	1.200	1	.479	98	2.503	.117
Tidak_Ada_Gratifikasi	4.286	1	.463	98	9.262	.003
Tidak_Ada_Pungli	4.040	1	.775	98	5.213	.025
Tidak_Ada_Percaloan	4.540	1	.572	98	7.942	.006

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Fig. 6: ANOVA Table

Analysis:
The magnitude or size of the F value affects the Sig value in the ANOVA Table. If the F value is larger, the Sig value will be smaller and vice versa. The variable that shows the most difference with other variables is the variable of serving interesting drinks with an F value of 13,739 and a Sig value of 0,000. While the variable that shows the most similarity with the other variables is the coffee variable that is in high demand with an F value of 2,503 and a Sig value of 0,117.

Number of Cases in each Cluster

Cluster	1	99.000
	2	1.000
Valid		100.000
Missing		.000

Fig. 7: Number of Cases in Each Cluster

Analysis:
Number of Cases in each Cluster is an output that indicates the number of respondents in a cluster. In cluster 1 the number of respondents was 99 and in cluster 2 the number of respondents was 1. The number of valid data is as many as 100 data and no data is missing.

2) Hierarchial Cluster

Case Processing Summary[a]

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
100	100.0	0	.0	100	100.0

a. Average Linkage (Between Groups)

Fig. 8: Case Processing Summary

Analysis:
The output above is a case processing summary. The function of this output is as a conclusion of the data analysis process which includes valid and missing data. Based on the table above, it can be seen that there is no missing data.

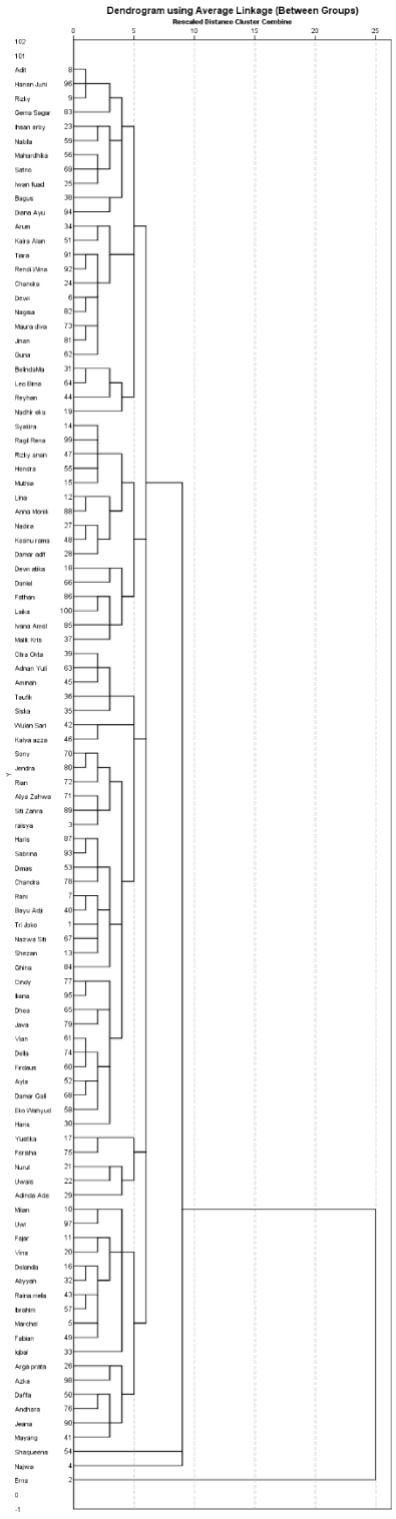


Fig. 9: Dendogram using Average Linkage

Analysis:
Based on the output of the dendogram above, it can be seen that the respondents are divided into 2 large clusters.

4.2. Discussion Analysis

This study was conducted with the aim of analyzing passenger satisfaction at the Purabaya Type A Terminal and grouping the variables into homogeneous groups. To obtain data was carried out by distributing questionnaires to passengers and to solve the problems in this study with the help of SPPS software and by cluster analysis methods which include K-means cluster tests and hierarchical cluster tests. The outputs generated from the Non-Hierarchy method are Initial Cluster Center, Iteration History, Final Cluster Center, ANOVA Table,

and Number of Each Cluster. Meanwhile, the outputs produced from the Hierarchy method are Case Processing Summary, and Dendogram Using Average Linkage.

1) Non-Hierarchial Cluster

a) Initial Cluster Center

In cluster 1 all variables which include adequate terminal facilities, information provided appropriately, adequate number of officers, easily accessible schedule, guaranteed terminal security, no discrimination, no officer fraud, no gratuities, no illegal levies, and no fraud have the same value, which is 5.00

In cluster 2 all variables including adequate terminal facilities, information provided appropriately, adequate number of officers, guaranteed terminal security, no discrimination, no gratuities, no illegal levies, and no fraud have the same value, which is 2.00. As for the variables that include the schedule, it is easily accessible and there is no fraud, the officer has a value of 3.00.

b) Iteration History

The output above is the Iteration History. The iterations carried out are as many as 2 iterations to achieve the optimum value. The shortest distance between cluster centers, which is 8,944.

c) Final Cluster Center

The output above is the Final Cluster Centers. The number of clusters formed is 2 clusters. It can be seen that all variables are included in cluster 1, namely the variables of adequate facilities, appropriate information, adequate number of officers, easily accessible schedules, guaranteed security, no discrimination, no fraud, no gratuities, no illegal levies, and no fraud. This is because all of these variables have higher values in cluster 1. Meanwhile, no variables go into cluster 2 because all variables have lower values in cluster 1.

d) ANOVA Table

The magnitude or size of the F value affects the Sig value in the ANOVA Table. If the F value is larger, the Sig value will be smaller and vice versa. The variable that shows the most difference with other variables is the variable of serving interesting drinks with an F value of 13,739 and a Sig value of 0,000. While the variable that shows the most similarity with the other variables is the coffee variable that is in high demand with an F value of 2,503 and a Sig value of 0,117.

e) Number of Cases in Each Cluster

Number of Cases in each Cluster is an output that indicates the number of respondents in a cluster. In cluster 1 the number of respondents was 99 and in cluster 2 the number of respondents was 1. The number of valid data is as many as 100 data and no data is missing.

2) Hierarchial Cluster

a) Case Processing Summary

The output above is a case processing summary. The function of this output is as a conclusion of the data analysis process which includes valid and missing data. Based on the table above, it can be seen that there is no missing data.

b) Dendogram using Average Linkage

Dendogram Based on the Dendogram, it can be seen that the respondents are divided into 2 large clusters. Cluster 1 consists of Muthia to Shaqueena. And cluster 2 consists of Milan to Erna. The cluster was then divided into smaller clusters based on the similarity of respondents' perceptions in filling out questionnaires regarding the level of passenger satisfaction with services at the Purabaya Type A Terminal

5. Conclusions and Suggestions

This research was conducted at the Purabaya Type A Terminal with data processing carried out using SPSS software from data in the form of questionnaire results. From this data, it will be processed using the cluster method which is divided into two types of methods, namely Non Hierarchial Cluster which produces output in the form of Initial Cluster Centers, Iteration History, Final Cluster Centers, ANOVA, and Number of Cased In Each Cluster and with the Hierarchial Cluster method which produces output in the form of Case Processing Summary and Dendogram Using Average Linkage. For Non-Hierarchial clusters in the Final Cluster Center, the number of clusters formed is 2 clusters. It can be seen that all variables are included in cluster 1, namely the variables of adequate facilities, appropriate information, adequate number of officers, easily accessible schedules, guaranteed security, no discrimination, no fraud, no gratuities, no illegal levies, and no fraud. This is because all of these variables have higher values in cluster 1. Meanwhile, no variables go into cluster 2 because all variables have lower values in cluster 1. Then iterations are needed 2 times with the closest distance from the cluster center of 8,944. For Hierarchial Cluster, the results of processing Cluster Membership data were obtained. Through analysis using the cluster method, valid data values of (N) 100 data with a value of 100%, and missing data as much as (N) 0 with a value of 0%. Based on the results of the data processing obtained, it is hoped that the Purabaya Type A Terminal can be better in the future. And for the next research, it is recommended that the number of respondents can be increased so that the research results can be even better, especially in terms of validity.

References

[1] Sudjana, Metode Statistika, Bandung: Tarsito, 2005

[2] Tajul, M, Learning SPSS Without Pain, Bangladesh: Asa Publications, 2020

[3] Tarigan, E.F., Nilmarito, S., Islamiyah, K., Darmana, A., & Suyanti, R.D. (2022). "Analisis Instrumen Tes Menggunakan Rasch Model dan Software SPSS 22.0." Jurnal Inovasi Pendidikan Kimia, 16(2). 92-96. http://dx.doi.org/10.15294/jipk.v16i2.30530

[4] M.M. Ramly, Sudarmin, B. Poerwanto. (2022). "Analisis Hierarchial Clustering Multiscale Bootstrap (Kasus: Indikator Kemiskinan di Provinsi Sulawesi Selatan Tahun 2020)." Journal of Statistics and Its Application on Teaching and Research, 4(3). 142-152. https://doi.org/10.35580/variansiunm26

[5] Joseph, et al, Multivariate Data Analysis, Upper Saddle River: Prentice Hall, 2009

[6] Natasya & Arie. (2021). "Analisis Cluster Kemiskinan Provinsi di Indonesia Tahun 2019 dengan Metode Partitoning dan Hierarki." Jurnal Sistem Komputer, 10(2). 101-109. http://dx.doi.org/10.34010/komputika.v10i2.4317

[7] A. Tifani, D.B.A. Pramana, B.N Sari. (2022). "Metode K-Means Clustering dalam Pengelompokkan Penjualan Produk Frozen Food." Jurnal Ilmiah Wahana, 8(15). 574-583. https://doi.org/10.5281/zenodo.7052276

[8]. Muhajir, N. Ismail, S. Syahreza. (2021). "Pemutakhiran Zona Musim (ZOM) Provinsi Aceh Menggunakan Data Blending Berbasis Non-Hirarki K-Means Cluster." Jurnal Ilmiah Fisika FMipa Universitas Lambung Mangkurat, 18(1). 35-41. http://dx.doi.org/10.20527/flux.v18i1.8746