APPLICATION OF DATA MINING ON SALES PRODUCT IN MATERNAL DISASTER DISTRO USING CLUSTERING METHOD

Dian Fitra N[1], Dian Dharmayanti[2]

[1,2] *Indonesian Computer University*
Jl. Dipatiukur 112-116, Bandung 40132, Indonesia
E-mail : dianfitranawangsari@gmail.com[1], diandmdr@yahoo.com[2]

ABSTRACT

Maternal Disater Distro is store that sells clothing and located in Bandung since 2003. The main products are sold in many Maternal Disaster is clothing. After conducting interviews, obtained party information distributions always make new products every month and delivered to outlets that have been scattered. Based on the survey questionnaire against the 14 outlets, obtained information that 11 of the 14 outlets stating the product that is sent to every outlet does not comply with the characteristics of the outlet. Judging from the results of a survey that has been done, can be drawn the conclusion that the delivered product does not comply with the characteristics of the different outlet so that the outlet had difficulty in selling the products of Maternal.

To get information of the characteristics of each outlet, then the segmentation is done against every outlet uses data mining and clustering method. The algorithm used is a Hierachical Agglomerative Clustering (AHC) where segmentation starts with presenting any data objects as one segment, then performed a calculation of the distance between the segments. After that, a segment that has a distance of terkedat are combined into a new segment. The solution to these problems is to make up the segmentation outlet based on sales data, so characteristic of each outlet can be seen.

Based on the test results it can be concluded that the segmentation of the outlet can facilitate Distribution to form party segmentation outlet to ease party distros in determining the right business strategy for each segment of outlets based on its characteristics.

Kata kunci : *outlet segmentation, data mining, clustering, Agglomerative Hierarchical Clustering, complete linkage*

1. INTRODUCTION

Maternal Disater Distro is store that sells clothing and located in Bandung since 2003. The main products sold in distributions Maternal Disaster is clothing. In addition to clothing, other supporting products such as bags, wallets, jackets, pants, and hats are sold in some outlets are scattered.

After conducting interviews with the Maternal Disaster Distro, obtained information that they always make new products every months and those product are being delivered to outlets that had been scattered. Based on questionnaire surveys to 14 outlets, obtained information that 11 of 14 outlets declared the product that being delivered to every outlets not in accordance with the needs of outlets. The mismatch happen because the product's type and size is different in each outlet, making each outlet has different requirements for the product. Seen from survey's results that was conducted, can be concluded that the delivered product was not match with the requirements and make the outlets troubled at selling Maternal's products.

To get information from every outlet, then the necessary sales data from every outlet in August 2016 to be analyzed. Methods that can be used in analyzing a lot is a method of data mining, that the methods used to see patterns or large data sets [1]. There are several methods that can be used depending on the objectives that will be achieved. One method that can be used is a method of clustering, i.e. the method of separating the data into a number of groups (clusters) according to the similarity of the characteristics of the data.

Based on test result that has been presented before, a possible solution is performed by forming segmentation outlet so that the characteristics of the product needs of each outlet can be seen. The characteristics of the outlet can be seen based on sales data has been done. Therefore, segmentation will be done based on sales data. Segmentation outlet can become a knowledge by analyzing the different characteristics or behaviors that require different products from each segmentasinya. So are expected to maximize the process of delivery of products according to the segmentation of the outlet that has been done.

*Thus Maternal Disaster need to use methods of data mining and clustering in the specify segmentation of data product sales outlets. Therefore from the explanations already described above, need to be made an application "**Application of Data Mining on Sales Product in Maternal Disaster Distro Using Clustering Method**".*

1.1 Data Mining

Data mining is an activity extract or mine knowledge from data or a large-sized. Other names which are often used and has the same meaning as data mining is knowledge from data mining, knowledge extraction, pattern analysis, data archaeology, data dredging.

Data mining seeks to solve problems by analyzing a set of data that already exist in the database. While the result of the process of data mining is knowledge discovery, information, patterns, or model [6].

1.2 Clustering

Clustering is the process of grouping objects based on information obtained from data describing relationships between objects by principles to maximize the similarity between members of one class and minimising the similarities between classes or clusters [7]. His goal was to find a cluster of quality in a decent time. Clustering in data mining is useful for finding the distribution pattern in a data set are useful for the process of data analysis. The similarity of the objects is usually obtained from the proximity of the attribute values that describes the data objects, while objects data is typically represented as a point in space is multidimensional.

With clustering, a densely populated areas can be identified, the patterns of the distribution as a whole and the interesting linkages between attribute-attribute data. In the data mining efforts focused on discovery methods for cluster on large databases effectively and efficiently.

Generally there is some category clustering algorithm known as follows:

a. method of Partition (Partition-based Clustering)
In this method, the user must specify the number of the desired partition and then k every contested data to be included on one of the partitions so that no data overlap and one data has only one cluster. For example: the K-Means algorithm, mixture modeling.

b. method of Hierarchy (Hierarchical Clustering)
Create a hierarchy in the form of a dendogram where the data will be stored in a hierarchy similar to the adjacent. This method is divided into two i.e. buttom-up that combines small clusters into larger clusters and top-down that break down large clusters into smaller clusters. Agglomerative Hierarchical Clustering For Example: (AHC), Divisive Hierarchical Clustering (DHC).

Clustering method in the important thing is to measure the similarities or dissimilarities into data to be grouped. Similaritas (the similarities) between two objects is a numerical measure of how similar two objects. Similaritas have a higher value if the objects are equal. Similaritas value range is [0,1]. Dissimilaritas (total dissimilarity) between two objects is a numerical measure of how two different objects. Dissimilaritas is worth lower if the objects are similar. Dissimilaritas minimum is 0 and the limits of it varies. Dissimiliratas is often called also with distance (distance) [8]. One way distance measurements is to use Euclidian distance (Euclidian Distance), which is expressed by equation (2.1) the following:

$$d' = \sqrt{\sum_{i=1}^{r}(xi - yi)^2} \qquad \text{Equation (2.1)}$$

where :
d = distance
r = the number of features in the vector
xi = the value of i to x feature
yi = the value of the feature y to i

1.3 Agglomeration Hierachical Clustering (AHC)

Grouping data using the agglomerative hierarchical clustering (AHC) starts with represent each object in the data as a single group, then do the calculation of distance (distance measure) between groups. After that the two groups who have the closest distance are combined into a new group. Then the distance between the new group with other groups is calculated using one of the methods of calculation of the distance between groups (single linkage, complete linkage, average linkage, etc.) [9] the process of calculation of distance and the merger of the two groups conducted in iterative. The following are some of the methods that can be used to calculate the distance between groups:

a. Single Linkage
Single linkage is also called minimum links, where the similaritas of the two clusters based on the closest two points from two different clusters. To search for single linkage can use equation (2.2) the following:

$$duv = \min\{duv\}, duv \in D \qquad \text{Equation (2.2)}$$

b. Complete Linkage
Complete linkage is also called the maximum link, in which the distance between two clusters is calculated from the furthest distance between members of a cluster with a second cluster. To search for complete linkage can use equations (2.3):

$$duv = \max\{duv\}, duv \in D \qquad \text{Equation (2.3)}$$

c. Average Linkage
Average linkage is also called group average, where the distance between two clusters is calculated from the average distance between the cluster members to each cluster. To find the average linkage can use equations (2.4):

$$duv = \text{average}\{duv\}, duv \in D \qquad \text{Persamaan (2.4)}$$

The following is the steps in the algorithm of Agglomerative Hierarchical Clustering (AHC) [10]:

1. *Present each outlet as one group, then calculate the distance (distance measure) between the groups.*
2. *Locate the two clusters that have the closest distance using complite linkage (found in equations (2.3).*
3. *Repeat steps one and two until the left one group (a collection of all outlets) and stops when the number of clusters are formed is equal to 1.*
4. *When the number of clusters amounted to 1, then the 3rd stage has been completed.*

Dendrogram cuts can be done based on agglomeration schedule. The number of clusters can be obtained by finding the difference koofisien the distance from each iteration.

1.4 Dendogram

Dendogram or tree structure is typically used to represent the process of clustering-based hierarchy, that shows how an object is grouped in each stride. Here is an example of a dendogram can be seen in Figure 1:

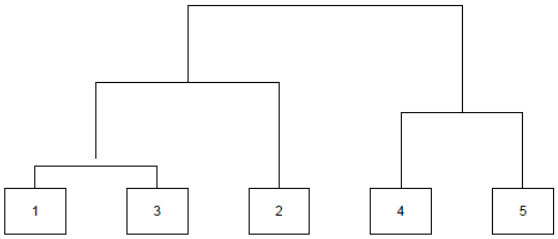


Figure 1. *Dendogram Based Clustering Results Hierarchy*

1.5 Agglomeratin Schedule

Agglomeration schedule is a numerical summary of cluster solutions resulting from the clustering process using hierarchical methods. [11] the following is an example of agglomeration schedule:

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	8	11	1.260	0	0	7
2	6	7	1.579	0	0	4
3	2	9	1.625	0	0	7
4	1	6	2.318	0	2	6
5	3	5	2.619	0	0	8
6	1	10	3.670	4	0	10
7	2	8	4.420	3	1	8
8	2	3	4.505	7	5	9
9	2	4	4.774	8	0	10
10	1	2	5.718	6	9	0

Figure 2. *Examples Of Agglomeration Schedule*

Agglomeration schedule table explanation:

1. *The stage is stages or iterations that have been made on the process of clustering using the hierarchical methods.*
2. *Cluster combined is a cluster which is joined on each stage.*
3. *Coficients is the distance of the two clusters that are incorporated in any stagenya are searchable by using the formula of the distance.*
4. *Stage cluster first appear is the stage when the combined cluster has merged with other clusters on a stage before. If clusters combined have never been combined at the previous stage, it will be filled with a value of 0.*
5. *The next stage is the stage next when the cluster will be combined with other clusters. [11].*

2. CONTENT OF RESEARCH

In this section will be discussed regarding the stages of the research done.

2.1 Business Understanding

Business goals of Maternal Disaster Distro is determining the right business strategy for each outlet. An important factor in order for the business purpose of Maternal Disaster Distro is achieved by knowing the characteristics of every outlet, so that the products delivered will be sold in accordance with the characteristics of the outlet.

Determination of target data mining consists of the goals and criteria of successful data mining. The following is the purpose and criteria for successful data mining:

1. *Data Mining Goals*
   The purpose of the application of data mining in the research is to establish segmentation outlet in Maternal Disaster Distro. This segmentation will be used by the promotion section in Maternal Disaster Distro as a consideration in order to determine the right business strategy for each outlet.
2. *Successful Criteria of Data Mining*
   The criteria of success in this research is able to assist in determining promotion section in Maternal Disaster Distro of segmentation outlet has the same characteristics in the product sales.

2.2 Data Understanding

Stage of understanding the data is the second tahapn which there are a number of tasks, including:

2.2.1 Collecting Early Data

The data used in the initial data collection is the sales data in August 2016 in distro Maternal Disaster originating from Microsoft Excel (.xlsx) who conducted the process of mining on an application. Here is a sample table sales data in August 2016:

MATERNAL DISASTER — LAPORAN DATA PENUALAN

Arena Experience - Bandung

No	ID Barang	Nama Artikel	Ukuran	Jenis	Sub Jenis	Warna	Harga	Jumlah	Dibayar
1	02607080101	SCR.04	AS	ACCESSORIES	bandana	black	Rp 70.000	1	Rp 70.000
2	02607081001	SCR.05	AS	ACCESSORIES	bandana	red	Rp 70.000	1	Rp 70.000
3	03207041201	BHOF	AS	ACCESSORIES	belt	brown	Rp 295.000	3	Rp 885.000
4	03407100102	FG-01	AS	ACCESSORIES	gloves	black	Rp 130.000	1	Rp 130.000
5	03407100101	GLV 01	AS	ACCESSORIES	gloves	black	Rp 160.000	1	Rp 160.000
6	02707140101	FOCKS	AS	ACCESSORIES	socks	black	Rp 60.000	2	Rp 120.000
7	02707140501	BOLOCK WHITE	AS	ACCESSORIES	socks	white	Rp 60.000	2	Rp 120.000
8	02907140102	THSOCK	AS	ACCESSORIES	socks	black	Rp 60.000	1	Rp 60.000
9	03407010101	LODGE	AS	ACCESSORIES	wallets	black	Rp 330.000	2	Rp 660.000
10	03303011701	GALAD	AS	BAG	backpack	olive green	Rp 310.000	1	Rp 310.000
11	02903011601	GOTHA	AS	BAG	backpack	dark green	Rp 290.000	4	Rp 1.160.000
12	02703020101	RAFHS	AS	BAG	sling bag	black	Rp 275.000	4	Rp 1.100.000
13	02703021601	DAGTF	AS	BAG	sling bag	dark green	Rp 275.000	3	Rp 825.000
14	02903020101	NILECA	AS	BAG	sling bag	black	Rp 290.000	2	Rp 580.000
15	02903020102	CALLER	AS	BAG	sling bag	black	Rp 290.000	1	Rp 290.000
16	03203060801	GLUTO	AS	BAG	waist bag	navy blue	Rp 225.000	2	Rp 450.000
17	02603060301	3AIWE	AS	BAG	waist bag	gray	Rp 205.000	1	Rp 205.000
18	02603060102	GATHER	AS	BAG	waist bag	black	Rp 205.000	1	Rp 205.000
19	02703060103	SCYV	AS	BAG	waist bag	black	Rp 210.000	1	Rp 210.000

Figure 1. *Transaction Data in August 2016*

2.2.2 Data Explanation

Initial data that has been collected will be explained at this stage. There are 10 on the data attribute, namely No, ID Barang, Nama Artikel, Ukuran, Jenis, Sub Jenis, Warna, Harga, Jumlah, Dibayar.

2.2.3 Data Exploration

Stages of data exploration can help the purpose of data mining. At this stage, carried out in-depth analysis against the data that has been retrieved. These activities are conducted to find out the characteristics and quality of the data. But at this stage of exploration data is not done because the data presented clearly without having to perform data exploration phase.

2.2.4 Data Quality Evaluation

Evaluation of the quality of the data is done against the number of sales based on the attribute jenis, sub jenis, warna, and ukuran. Following is the result of the evaluation of attribute-the attribute.

1. *The exploration phase is not necessary in this case because the data presented clearly without having to perform data exploration phase. Data exploration conducted to know the quality of the data.*
2. *Missing value might be found on the sales data. But on sales data missing value was not found.*

2.3 Data Preparation

At this stage of processing will be performed against the data that has been retrieved (data preprocessing). Data preprocessing need to be done to improve the quality of the data to be used in the modeling process, there are some tasks in this stage, namely:

2.3.1 Data Selection

The selection of the data was conducted to determine the data that will be used in clustering or process modeling stage. The selection of such data include the selection of attributes and record that will exist in the system. There are four attributes that will be selected for the modeling stage i.e. jenis, sub jenis, warna, ukuran, and jumlah. Attributes and record were selected because in accordance with the business purpose of the Maternal Disaster Distro.

2.3.2 Data Cleaning

The process of data cleansing is done to improve the quality of the data obtained. Data cleansing refers to the evaluation of the quality of the data has been previously performed. If there is a missing value, then the value of the missing will be populated with a default value of 0. Not done an analysis of an outlier since the algorithms used are not affected by the value of an outlier.

2.3.3 Data Arrangement

This process is done on the construction of a new set of data or produce derived attributes through the transformation of data modeling needs. Analysis on the transformation of the time this involves sales data in August 2016, lowers attributes required at the stage of modeling and discard the attribute is not required for the modeling stage. The selected attribute in the compilation of this data is type, sub type, color, and size of the combined use of the sign of the underscore and given the name sales transaction Data. Following the results of sales transaction data is as follows:

ID OUTLET	ACCESSORIES_BANDANA_BLACK_AS	ACCESSORIES_BANDANA_NAVY BLUE_AS	ACCESSORIES_BANDANA_RED_AS	ACCESSORIES_BANDANA_WHITE_AS	ACCESSORIES_BELT_BLACK_AS	ACCESSORIES_BELT_BROWN_AS
MT01	13	1	0	0	0	0
MT02	0	0	0	2	0	2
MT03	0	0	0	1	0	0
MT04	2	0	0	1	0	0
MT05	0	0	0	0	0	0
MT06	12	0	1	2	0	0
MT07	2	0	0	1	0	0
MT08	0	0	0	4	0	0
MT09	2	0	0	0	2	0
MT10	2	0	0	1	0	0
MT11	0	0	0	2	0	2
MT12	4	0	0	1	0	1
MT13	4	0	0	0	0	0
MT14	0	0	1	0	0	0

FIgure 2. *Data Arrangement*

2.3.4 Data Integration

At this stage done merging two or more tables that have different information about the same object into a new set of data. However, at this stage, integrity is not done, because the necessary data do not need to perform the merge.

2.4 Modeling

The modeling stage was the main activity of problem solving process of data mining. Data is already arranged will be processed using a method so that it can generate the model. Activities at the stage of modelling include modelling techniques are used, the mechanism of the model used, as well as the analysis of test model.

2.4.1 Modeling Techniques

Modeling techniques that will be used i.e. using herarchical clustering, and the model that will be used is an Agglomerative algorithm Herarchical Clustering (AHC). The method is used because it can make up the segmentation outlet is naturally based on similarity of attribute values in the data. In addition, the number of outlets to be formed segmentation is not yet known. Using that method, the desired amount of segmentation may be determined after the model is formed. The results of the segmentation outlet will be presented on a dendogram.

2.4.2 Modeling Mechanism

At this stage of the mechanism explained a model that will be used as well as the initial assumptions of the model. Agglomerative hierarchical clustering algorithm model has the following stages:

5. *Present each outlet as one group, then calculate the distance (distance measure) between the groups.*
6. *Search two clusters that have the closest distance using complite linkage (found in equations (2.3)).*
7. *Repeat steps one and two until the left one group (a collection of all outlets) and stops when the number of clusters are formed is equal to 1.*
8. *When the number of clusters amounted to 1, then the 3rd stage has been completed.*

2.4.3 The Use of Modeling Techniques

At this stage, researchers will apply the modelling methods have been described previously. The following is the process of the implementation of Agglomerative Hierarchical Clustering method on data that has already been prepared.

Step-1

Initialization initial cluster is as much as 14 clusters, because the data (n) = cluster (n), then calculate distances using Euclidean Distance is calculated using the following equation:

$$dist = \sqrt{\sum_{i=1}^{r}(xi - yi)^2} \qquad \text{equation (2.1)}$$

Each cluster will be calculated distance himself with other clusters. The following is the process of the calculation of the distance between clusters. Here is an example of the calculation of the distance by ukuran:

Dman(1,2) =

$$\sqrt{\begin{array}{c}(13-0)^2+(0-0)^2+(0-0)^2+(5-2)^2+(5-0)^2+(3-0)^2+(0-4)^2+(4-11)^2+\\(0-0)^2+(5-2)^2+(0-0)^2+(7-9)^2+(0-0)^2+(0-0)^2+2-0)^2+(17-31)^2+\\(0-7)^2+(1-5)^2+(0-4)^2+(0-2)^2+(0-0)^2+(7-11)^2+(11-15)^2+(8-11)^2+\\(4-13)^2+(1-2)^2+(5-24)^2+(18-17)^2+(30-20)^2+(2-10)^2+(7-4)^2+(1-10)^2+\\(3-8)^2+(3-5)^2+(0-2)^2+(1-2)^2+(1-0)^2+(2-0)^2+(0-0)^2+(0-0)^2+\\(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2+\\(0-0)^2+(0-0)^2+(0-0)^2+(1-1)^2+(1-0)^2+(0-0)^2+(2-0)^2+(0-0)^2+\\(1-1)^2+(0-0)^2+(0-1)^2+(0-0)^2+(0-0)^2+(0-0)^2+(0-1)^2+(6-1)^2+\\(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2+\\(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2\end{array}}$$

= 40,78

Calculation by using EuclideanDistance equations made against all data, so the distance matrix will be produced. For more complete the distance matrix can be seen in Appendix F.

Step-2

At this stage will be calculated the distance between two clusters by using complete linkage (equation 2.3). using this method will look for the closest spacing of two clusters and cluster merging with maximum distance and generate new cluster. The following example calculation of Euclidean Distance matrix.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	40,78	36,10	38,00	31,84	28,07	33,30	32,63	35,04	33,72	35,57	40,46	34,77	31,42
2	40,78	0	29,43	26,93	33,12	36,43	25,61	43,31	24,68	36,77	24,08	28,95	27,89	32,22
3	36,10	29,43	0	24,72	29,34	27,00	20,40	36,41	23,22	31,84	23,87	24,17	19,03	26,57
4	38,00	26,93	24,72	0	34,15	32,34	25,98	42,74	28,84	34,34	28,62	29,27	28,09	29,38
5	31,84	33,12	29,34	34,15	0	26,53	29,51	32,30	28,21	35,76	30,08	29,00	26,25	27,87
6	28,07	36,43	27,00	32,34	26,53	0	26,36	35,62	28,25	34,83	29,78	28,37	24,76	27,77
7	33,30	25,61	20,40	25,98	29,51	26,36	0	36,88	20,22	26,34	21,49	25,57	20,45	22,89
8	32,63	43,31	36,41	42,74	32,30	35,62	36,88	0	37,75	39,45	39,14	40,15	37,92	32,98
9	35,04	24,68	23,22	28,84	28,21	28,25	20,22	37,75	0	31,42	20,57	23,43	20,76	25,20
10	33,72	36,77	31,84	34,34	35,76	34,83	26,34	39,45	31,42	0	30,50	38,94	32,68	25,85
11	35,57	24,08	23,87	28,62	30,08	29,78	21,49	39,14	20,57	30,50	0	25,30	23,24	24,74
12	40,46	28,95	24,17	29,27	29,00	28,37	25,57	40,15	23,43	38,94	25,30	0	20,90	29,55
13	34,77	27,89	19,03	28,09	26,25	24,76	20,45	37,92	20,76	32,68	23,24	20,90	0	23,52
14	31,42	32,22	26,57	29,38	27,87	27,77	22,89	32,98	25,20	25,85	24,74	29,55	23,52	0

Figure 3. *Euclidean Distance matrix*

Min(Dman) = min (D(3,13) = 19,03. Of all existing cluster, cluster 3 and 13 selected because it has the most value for a minimum of another cluster. Calculate the distance of 3 and 13 with other clusters are left. With the following example:

D(3,13) 1 = maks{d(3,1) d(13,1)} = maks{(36,10) (34,77)} = 36,10. After membandingan two values (36.10) and (34.77) selected the maximum value. After the calculation of distances for all clusters, d (3.13) will form a new cluster, so the cluster 3 and 13 are removed. With the following example:

	(3.13)	1	2	4	5	6	7	8	9	10	11	12	14
(3.13)	0	36,10	29,43	28,09	29,34	27,00	20,45	37,92	23,22	32,68	23,87	24,17	26,57
1	36,10	0	40,78	38,00	31,84	28,07	33,30	32,63	35,04	33,72	35,57	40,46	31,42
2	29,43	40,78	0	26,93	33,12	36,43	25,61	43,31	24,68	36,77	24,08	28,95	32,22
4	28,09	38,00	26,93	0	34,15	32,34	25,98	42,74	28,84	34,34	28,62	29,27	29,38
5	29,34	31,84	33,12	34,15	0	26,53	29,51	32,30	28,21	35,76	30,08	29,00	27,87
6	27,00	28,07	36,43	32,34	26,53	0	26,36	35,62	28,25	34,83	29,78	28,37	27,77
7	20,45	33,30	25,61	25,98	29,51	26,36	0	36,88	20,22	26,34	21,49	25,57	22,89
8	37,92	32,63	43,31	42,74	32,30	35,62	36,88	0	37,75	39,45	39,14	40,15	32,98
9	23,22	35,04	24,68	28,84	28,21	28,25	20,22	37,75	0	31,42	20,57	23,43	25,20
10	32,68	33,72	36,77	34,34	35,76	34,83	26,34	39,45	31,42	0	30,50	38,94	25,85
11	23,87	35,57	24,08	28,62	30,08	29,78	21,49	39,14	20,57	30,50	0	25,30	24,74
12	24,17	40,46	28,95	29,27	29,00	28,37	25,57	40,15	23,43	38,94	25,30	0	29,55
14	26,57	31,42	32,22	29,38	27,87	27,77	22,89	32,98	25,20	25,85	24,74	29,55	0

Figure 4. *Euclidean Distance matrix iteration-1*

Step-3

Repeat steps one and two until the left one group (a collection of all outlets) and will stop when the number of clusters are formed is equal to 1.

Step-4

When the number of clusters amounted to 1, then the phase 3 has been completed and the final result of looping based on size can be seen in Appendix d. here's a dendogram of grouping results. as shown in Figure 5. Final Results Of The Dendogram.

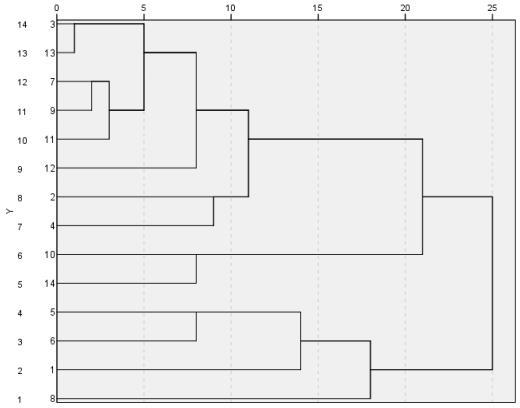


Figure 5. *Final Results Of The Dendogram*

A dendogram based have been established, the right level or number of clusters would best be determined by looking at the difference between the distance from each clusteryang is formed. To know the best number of clusters, one of which with a use agglomeration schedule. Agglomeration schedule can be seen in Figure 6.

Iterasi	Kombinasi Cluster		Koefisien	Iterasi Kombinasi Cluster Sebelumnya		Iterasi Selanjutnya
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	13	19,03	0	0	4
2	7	9	20,22	0	0	3
3	7	11	21,49	2	0	4
4	3	7	23,87	1	3	5
5	3	12	25,57	4	0	9
6	10	14	25,85	0	0	12
7	5	6	26,53	0	0	10
8	2	4	26,93	0	0	9
9	2	3	29,43	8	5	12
10	1	5	31,84	0	7	11
11	1	8	35,62	10	0	13
12	2	10	38,94	9	6	13
13	1	2	43,31	11	12	0

Figure 6. *Agglomeration Schedule*

In the first stage, a cluster is a cluster of 3 and coupled cluster 13, because it has the smallest distance i.e. 19.03. The cluster that has been incorporated in the second iteration will be consolidated back in the third and fourth iteration iteration will be consolidated back in the fifth iteration. On the 9th iteration, clusters that have been incorporated in cluster 5 and cluster 8 will be merged back. The cluster that will be generated will appear on the next iteration, so on to 13 iterations.

The number of files the best can be seen from the difference in's of each iteration. The difference of the coefficients based on segmentation ukuran can be seen in Figure 7.

Jumlah Cluster	Tahap Terakhir Agglomeration	Koefisien Tahap Ini	Selisih
2	43,31	38,94	4,37
3	38,94	35,62	3,32
4	35,62	31,84	3,78
5	31,84	29,43	2,41
6	29,43	26,93	2,5
7	26,93	26,53	0,4
8	26,53	25,85	0,68
9	25,85	25,57	0,28
10	25,57	23,87	1,7
11	23,87	21,49	2,38
12	21,49	20,22	1,27
13	20,22	19,03	1,19

Figure 7. *Difference In Coefficient Of Grouping*

The biggest difference that is the best number of clusters so 4.37 is 2 cluster. Then the dendogram results grouping based on the size of the cuts will be carried out as shown in Figure 8.

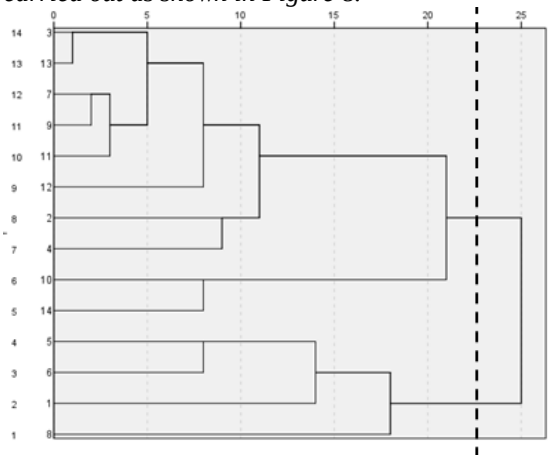


Figure 8. *Cutting A Dendogram*

The end result of this segmentation, formed two clusters with each group can be seen in table 1 and table 2.

Table 1. *Cluster 1*

ID Outlet	Nama Outlet	Alamat
MT05	Lawless	Jakarta Selatan
MT06	Distroclub	Bogor
MT01	Flagship Store	Bandung
MT08	Omuniuum	Bandung

Table 2. *Cluster 2*

ID Outlet	Nama Outlet	Alamat
MT03	Family Affair	Bandung
MT13	Bloop Store	Semarang
MT07	Boredom	Sumedang
MT09	Treshold	Bandung
MT11	Inspired	Malang
MT12	Arena	Bali
MT02	Sevensoul	Yogyakarta
MT04	Arena Experience	Bandung
MT10	Redneck Store	Sukabumi
MT14	District Store	Jakarta Pusat

From the results of the segmentation outlet has been done based on product sales, the highest sales results obtained from each cluster are as follows:

Table 3. *Highest Sales*

Cluster	***Jenis, Sub Jenis, dan Ukuran***	***Jumlah***
Cluster 1	*SHIRT_SHIRT RAGLAN_WHITE_XXL*	*71*
Cluster 2	*SHIRT_SHIRT BASIC SHORT_DARK BROWN_L*	*150*

From the highest sales results, product recommendations can be generated for each cluster can be seen in table 4 below:

Table 4. *Product Recommendations Per Cluster*

Cluster	***Rekomendasi Produk***
Cluster 1	*Jenis produk Shirt dengan sub jenis produk Shirt Raglan dengan warna White dan ukuran XXL*
Cluster 2	*Jenis produk Shirt dengan sub jenis produk Shirt Basic Short dengan warna Dark Brown dan ukuran Large*

3. CLOSING

Results from research that has been done, the conclusion that the application of the formation of ssegmentasi outlet that has been built up can help manager promotion to form the appropriate outlet segmentation with the similarity of product sales so referable to distribution and business strategies appropriate to the characteristics of each outlet.

There are some suggestions that can be made for application development the establishment of segmentation outlet, among others:

1. *Add an attribute that is used in modelling, so that the results of the segmentation more maximum in determining the characteristics of the outlet.*
2. *Algorithms used can be expanded using other algorithms, so you can compare the results with segmentation algorithms complete linkage.*
3. *The methods used in the calculation of distances can be developed again by using the method of Manhattan, because with this method of calculation is more easily understood when compared to euclidean distance search method.*

BIBLIOGRAPHY

[1] S. B, Data Mining Teknik Pengumpulan Data Untuk Keperluan Bisnis, Yogyakarta: Graha Ilmu, 2007.

[2] P. C. J. C. R. K. T. K. T. R. C. S. dan R. W. , "CRISP-DM 1.0 Step-by-step data mining guide," 2000.

[3] Prof. Dr. Suryana, M.Si., "Metodologi Penelitian Model Praktis Penelitian Kuantitatif dan Kualitatif"

[4] Prasetyo, E., " Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab", Yogyakarta: Andi Yogyakarta, 2012.

[5] R. Ramakrishnan dan J. Gehrke, Database Management System Third Edition, Singapore, 2003.

[6] B. Susanto, "Data Preprocessing," Versi 1.2, 2013.

[7] P. Berkhin, "Survey of Clustering Data Mining Techniques".

[8] F. A. Hermawati, Data Mining, Yogyakarta: Andi Yogyakarta, 2013.

[9] P. Berkhin, "Survey of Clustering Data Mining Techniques".

[10] R. Xu dan D. Wunsch, Clustering.

[11] M. Caccam dan J. Refran, "Cluster Analysis".

[12] A.S. Rosa dan M. Shalahuddin, Modul Pembelajaran Rekayasa Perangkat Lunak (Terstruktur dan Berorientasi Objek), Bandung: Modula, 2013.

[13] Ir. Supardi, Y., "Semua Bisa Menjadi Programming Java Basic Programming", Jakarta: Elex Media,2010

[14] Simarmata, Janner., "Rekayasa Perangkat Lunak", Yogyakarta : Andi, 2010