IMPLEMENTATION OF DATA MINING ON BREAD SALES USING CLUSTERING METHOD IN ROTI RIKI

Muhamad Sandy Hasanudin[1], Dian Dharmayanti[2]

[1,2] Universitas Komputer Indonesia
Jl. Dipatiukur 112-116, Bandung 40132, Indonesia
E-mail : sandy.hasanudin@yahoo.com[1], diandmdr@yahoo.com[2]

ABSTRACT

Roti Riki is a company who produces any of processed flour. On the daily bussiness, Roti Riki produces any processed flour which is sweet bread with various kind of jam. Roti Riki do not distributes to end user or retailer directly. Their distributor who sell to them. Distributor pick up the products from warehouse directly in Roti Riki. The Distributor can't decide what they want to pick up to their truck. All of them just can take whatever that available in warehouse. Those processes impact of negativeness to sales in Roti Riki. It happens because distribution of various kind of product mismatches with distributor needs. As a result is some distributors return some products.

Based on problems that explained before, the solution for solving the problems is form distributor groups based on their transaction data. The result of that clusterization is will be new knowledge with analyzing various character, perhaps needed different kind of product or different sales promotion for each distributor. Clusterization of distributors can build using clustering method in data mining. One of algorithm who can do that is Agglomerative Hierarchical Clustering (AHC). AHC Algorithm have capability to build clusters based on proximity of distances between datas.

Based on the test, the conclusion is built application can help sales division to sell and bussiness strategies rightly for any of cluster based on the result of analyzed before.

Keyword : distributor clusterization, data mining, clustering.

1. INTRODUCTION

Riki is a bread business units engaged in the production of refined flour. In its operational activities, bread flour is processed producing Riki i.e. sweet bread fill with various variants such content meses, mocca, Brown, kopyor, blueberries and the bread matter producing Riki cake with green beans pia, pia Brown. Types of bread produced namely bottled sweet bread that is often found in the stalls on the roadside. The bread does not distribute Riki directly against seller or retail stalls. The Distributors which will supply to the stalls that serve customers directly. Each product taken by distributors, bread Riki usually records the data transaction along with the number of products that are loaded into the database.

After doing an interview with Chairman of the production section, obtained information that the Bread is sometimes added content variant Riki flavor on its products. Usually, bread making innovations to reproduce Riki option for distributors to expand market share. Innovations that are done like adding content with a sense of new variants and also increase the size of a bread course offered at a price markedly. Unfortunately there is always new content variants with certain size produced it sold over the counter. When the bread has been distributed in a specified period of not selling even stale, then the distributor-distributors usually return products that are not sold to the factory Riki bread. Party Marketing Riki Bread want to reduce the potential returns on minimal possible variants and offer the products according to the needs of distributors. Because each distributor has a need for different products.

Based on the results of observation of the problem that has been presented before, a possible solution is done is form a group of distributors so that the characteristics of every product needs a distributor can be known [1]. Needs of each distributor can be seen based on transaction data that has already been done. Thus segmentation will be done based on-

right data transactions.

The observations from the data of the transaction turns out each data has a pattern of data and have the links indicates that there are recurring events. To get such information more quickly and efficiently help needed an information technology, in this case i.e. Data Mining [2]. Data Mining is the process of finding a meaningful relationship, patterns and trends by examining in great gobs of data stored in storage by using pattern recognition techniques such as mathematical and statistical techniques [3]. In data mining

There are several methods i.e. prediction, classification, clustering, and associations. One such method is a method of clustering, i.e. the method of grouping a number of data or objects into clusters (groups) in the cluster so that each will contain data

as possible and different from other cluster object in [4].

Based on problems that have been presented before, this will be built on the research of application of Data Mining to apply methods of Clustering is useful to segment the distributors in Bread useful information as Riki for the recommendations offer products to distributors.

Analysis of the problem which is done is viewed and identify problems or obstacles in the research conducted. Required the existence of a problem formulation issue through analysis of pre-set. Based on the observations, found the problem as follows:

1. marketing Parties require knowledge about the characteristics of each distributor, so that the delivered products according to the needs of distributors based on transaction data.

2. the marketing Party Bread Riki need to apply different sales strategies against a different distributor characteristics based on the data of the transaction.

2. CONTENT OF RESEARCH

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

2.1 Bussiness Understanding

Business goals of Bread is bread producers Riki packaging that can serve the distributor-distributor according to their needs. As for being an important factor in order for the business purposes of bread this is achieved is by Riki provides and distributes products that comply with the request and the characteristics of the distributor.

The target of the application of data mining in sales of bread this is Riki to classify distributor-distributor based on transaction data.

The assessment of the situation is the writing of facts that exist in the place of research in detail. Those facts include availability of resources, the necessary requirements, and the benefits that will be gained from the research of data mining. Following is the result of the assessment of the situation in the bread of Riki.

a) Resource Availability

On this activity recorded the various resources available for research needs. These resources include hardware, software, data, and personnel. The resources available in this research for Riki Bread is as follows:

1. The availability of Hardware and software

Below is a table outlining the availability of hardware in Bread Riki to be used to run the software required for the research.

Table 1 Availability of Hardware

No	Hardware	Spesification
1	*Processor*	Intel Core i3 2102 3.0 GHz
2	*VGA*	Intel HD 3000 1024 MB
3	*Memory*	RAM 2GB
4	*Harddisk*	500 GB
5	*Monitor*	LCD 14 *inch*
6	*Keyboard*	*Standard Port* PS2
7	*Mouse*	*Standard Port* USB

While table 2 is software available that will be used to run applications processing data mining, building a system analysis results, and make research reports.

Table 2 Availability of Software

No	Software	Spesification
1	Operating System	Microsoft Windows 7
2	Program	a. Microsoft Office 2013 b. Net Framework 3.5 c. Xampp d. R Tools

1. Data availability

The available data for this research is the data of the transaction. Such data is collected and processed by the head of marketing. For the purposes of research, researchers have received permission from the head of marketing section to access and manipulate the data.

2. Availability of Personnel

The available personnel comes from internal and external Bread Riki. In this study, head of Marketing of bread Riki is the only personnel associated with it. Head of Marketing of bread that provides the data for Riki analyzed research on data mining.

b the necessary Requirements)

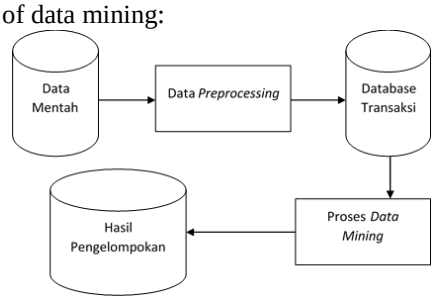
The needs that are required at this time, research is data distributor and data transactions during the month of may 2016 in Bread Riki.

c) benefit from research data mining.

A benefit of this data mining research is the existence of information and new knowledge for the marketing of bread Riki characteristic in this group of distributors based on transaction data. So that the information can be used as one of the

recommendations to make the right business strategy against distributors in Bread Riki.

Project planning stages that are traversed in order to reach the target goal of data mining. The following stages of project planning to achieve the target of data mining:



Picture 1 Project Plan

The raw data obtained from data transactions during the month of may 2016 in Bread Riki. Raw data for input into the system. The system performed against the data Preprocessing, then into a new dataset and entered into a database storage. Data that resides in the database then conducted the process of mining which then produce results grouping.

2.2 Data Understanding

Data understanding phase is the second stage of which there are a number of tasks, including the following:

2.2.1 Collect Data

The data in the research is the transaction data in a formatted Riki Bread Microsoft excel files (.xlsx) who conducted the process of mining on an application. Here is some initial data in the form of transaction data in may 2016.

ID	Nama	Alamat	Roti Manis 1000					Kue Pia	Roti Manis 2000				Sisa Roti Manis 1000					Sisa Kue Pia		Sisa Roti Manis 2000			
			Meses	Mocca	Coklat	Blueberry	Kopyor	Kacang Ijo	Meses	Kombinasi	Coklat	Kopyor	Meses	Mocca	Coklat	Blueberry	Kopyor	Kacang Ijo	Coklat	Meses	Kombinasi	Coklat	Kopyor
RRD-01	Toni Atmajaya	Purwakarta	7490	5830	4590	6550	4215	6400	2640	3320	3460	4170	3230	0	0	0	0	0	0	0	0	0	0	0
RRD-02	Asep Rachman	Ranca Ekek	3210	2750	4550	330	350	1330	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RRD-03	Aris Winarso	Baile Endah	2230	2230	3260	1050	1010	2260	860	585	820	865	410	0	0	0	5	0	5	0	0	0	0	0
RRD-04	Goyon Winarno	Cicalengka	3100	2535	3900	570	810	2330	550	440	290	685	190	289	320	422	15	35	267	65	0	0	97	0
RRD-05	Abas Yahya	Majalaya	6970	6040	4380	5590	4120	8470	2990	2553	2340	2900	2235	0	0	0	0	0	0	0	0	0	0	0
RRD-06	Mamo Surato	Majalaya	4670	3680	5500	1420	1800	3410	840	640	455	630	180	20	50	15	0	55	30	0	0	0	15	15
RRD-07	Aseng Sanjaya	Cibiru	3890	4230	5580	1700	1980	3420	870	825	200	885	290	24	124	0	0	0	32	0	0	0	0	0
RRD-08	Tursino Widagdo	Majalaya	3710	3500	5130	1800	1120	3870	1405	1215	800	1430	720	7	5	0	13	1	0	0	0	10	0	14
RRD-09	Imas Darmaji	Majalaya	3615	3340	3920	2740	2285	2925	2075	2115	1925	2255	2035	0	0	0	0	0	0	0	0	0	0	0
RRD-10	Ferdi Septian	Garut	3630	3080	4470	1675	1310	3255	1170	870	930	1225	550	323	130	199	66	44	225	24	15	13	11	0
RRD-11	Saprol Mahmud	Majalaya	4625	4585	5910	1445	1480	4015	810	720	475	860	365	0	47	0	0	0	28	0	35	0	0	0
RRD-12	Toyo Faturakhmat	Tasik	10150	11650	6290	7730	5170	7820	3585	4540	4755	5275	4530	10	5	0	5	0	22	0	0	0	0	8
RRD-13	Gilang Rambu	Garut	3980	3295	5350	1400	1770	3515	1440	830	920	1045	635	89	0	82	20	0	35	33	0	26	15	0
RRD-14	Sujoto	Sumedang	4960	3745	6345	1965	1710	4525	1870	1585	1210	1695	850	123	330	162	40	0	345	220	0	20	0	0
RRD-15	Atam Gunardi	Karawang	7530	8780	5740	9660	5850	9040	5160	4540	4070	5030	4100	0	0	0	0	0	0	0	0	0	0	0
RRD-16	Wasko Satrio	Lembang	10170	11600	4355	7805	4770	7770	2992	3150	2495	3650	2420	221	356	230	155	0	431	0	0	189	224	52
RRD-17	Yono Fathah	Sumedang	4880	9850	6090	2430	2285	5210	1795	1255	1190	2040	925	122	0	114	0	0	123	0	15	0	0	76
RRD-18	Lusiyuno Hadi Kus	Ranca Ekek	3180	2025	3850	1175	1330	2340	640	425	210	580	305	65	22	221	287	188	0	6	0	3	7	21
RRD-19	Otong Sirait	Sumedang	3870	3200	4640	1530	1790	3905	1420	1095	965	1475	640	0	0	0	0	0	0	0	0	0	0	0
RRD-20	Timo Wijaya	Cilacap	2900	3200	3110	1450	1250	2270	950	630	530	590	430	0	0	0	0	0	0	0	0	0	0	0

Figure 2. Transaction Data in Mei2016

2.2.2 Data Description

Initial data that has been collected will be explained at this stage. There are 25 data attribute, i.e. ID, Nama, Alamat, Roti Manis 1000 Meses, Roti Manis 1000 Mocca, Roti Manis 1000 Coklat, Roti Manis 1000 Kopyor, Roti Manis 1000 Blueberry, Kue Pia Coklat, Kue Pia Kacang Ijo, Roti Manis 2000 Meses, Roti Manis 2000 Kombinasi, Roti Manis 2000 Coklat, Roti Manis 2000 Kopyor, Sisa Roti Manis 1000 Meses, Sisa Roti Manis 1000 Mocca, Sisa Roti Manis 1000 Coklat, Sisa Roti Manis 1000 Kopyor, Sisa Roti Manis 1000 Blueberry, Sisa Kue Pia Coklat, Sisa Kue Pia Kacang Ijo, Sisa Roti Manis 2000 Meses, Sisa Roti Manis 2000 Kombinasi, Sisa Roti Manis 2000 Coklat, Sisa Roti Manis 2000 Kopyor

Table 3 Description of Attributes of Data Transactions

Description		
Function	To find out which products are purchased by a distributor as well as a number of product returns	
Attribute	ID	ID Distributor
	Nama	The Name of owner of distributor
	Alamat	Address of Distributor
	Roti Manis 1000 Meses, Roti Manis 1000 Mocca, Roti Manis 1000 Coklat, Roti Manis 1000 Kopyor, Roti Manis 1000 Blueberry, Kue Pia Coklat, Kue Pia Kacang Ijo, Roti Manis 2000 Meses, Roti Manis 2000 Kombinasi, Roti Manis 2000 Coklat, Roti Manis 2000 Kopyor, Sisa Roti Manis 1000 Meses, Sisa Roti Manis 1000 Mocca, Sisa Roti Manis 1000 Coklat, Sisa Roti Manis 1000 Kopyor, Sisa Roti Manis 1000 Blueberry, Sisa Kue Pia	Kind of Products

Description		
	Coklat, Sisa Kue Pia Kacang Ijo, Sisa Roti Manis 2000 Meses, Sisa Roti Manis 2000 Kombinasi, Sisa Roti Manis 2000 Coklat, Sisa Roti Manis 2000 Kopyor	

2.2.3 Data Exploration

Stages of exploration data 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 Verify Data Quality

Evaluation of the quality of the data is done against the number of sales based on the attribute type, sub type, color, and size. 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 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 Select Data

Select data 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 23 attributes that will be selected for the modeling phase i.e. ID, Roti Manis 1000 Meses, Roti Manis 1000 Mocca, Roti Manis 1000 Coklat, Roti Manis 1000 Kopyor, Roti Manis 1000 Blueberry, Kue Pia Coklat, Kue Pia Kacang Ijo, Roti Manis 2000 Meses, Roti Manis 2000 Kombinasi, Roti Manis 2000 Coklat, Roti Manis 2000 Kopyor, Sisa Roti Manis 1000 Meses, Sisa Roti Manis 1000 Mocca, Sisa Roti Manis 1000 Coklat, Sisa Roti Manis 1000 Kopyor, Sisa Roti Manis 1000 Blueberry, Sisa Kue Pia Coklat, Sisa Kue Pia Kacang Ijo, Sisa Roti Manis 2000 Meses, Sisa Roti Manis 2000 Kombinasi, Sisa Roti Manis 2000 Coklat, Sisa Roti Manis 2000 Kopyor.

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 Construct Data

This process is done on the development of a new data set or produce derived attributes through the transformation of data modeling needs. Lowers attribute-the required attributes in the modeling and discard the attribute is not required for the modeling stage. Following the results of the preparation of the data:

ID	Roti Manis 1000					Kue Pia	Roti Manis 2000			Sisa Roti Manis 1000					Sisa Kue Pia	Sisa Roti Manis 2000			
	Meses	Mocca	Coklat	Blueberry	Kopyor	Kacang Ijo	Coklat	Meses	Kombinasi	Coklat	Kopyor	Meses	Mocca	Coklat	Blueberry	Kopyor	Kacang Ijo	Coklat	Meses	Kombinasi	Coklat	Kopyor
RRD-01	7490	5830	4590	6502	4215	6400	2640	3320	3460	4170	3230	0	0	0	0	0	0	0	0	0	0	0
RRD-02	3210	2750	4520	300	350	1330	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RRD-03	2230	2230	3260	1050	1010	2260	860	585	820	865	410	0	0	0	5	0	5	0	0	0	0	0
RRD-04	3100	2535	3900	570	810	2330	550	440	290	685	190	289	320	422	15	35	267	65	0	0	37	0
RRD-05	6970	6040	4380	5390	4120	8470	2990	2535	2340	2900	2235	0	0	0	0	0	0	0	0	0	0	0
RRD-06	4670	3680	5500	1420	1800	3410	840	640	455	630	180	20	50	15	0	55	30	0	0	0	15	15
RRD-07	3890	4230	5580	1700	1590	3420	870	825	200	885	290	24	124	0	0	0	32	0	0	0	0	0
RRD-08	3710	3500	5130	1800	1120	3870	1405	1215	800	1430	720	7	5	0	13	1	0	0	0	10	0	14
RRD-09	3615	3340	3920	2740	2285	2925	2075	2115	1925	2255	2010	0	0	0	0	0	0	0	0	0	0	0
RRD-10	3630	3080	4470	1875	1310	3255	1170	870	930	1225	550	323	130	199	66	44	225	24	15	13	11	0
RRD-11	4625	4585	5910	1445	1480	4015	810	720	475	860	360	0	47	0	0	0	28	0	55	0	0	0
RRD-12	10150	11850	6260	7730	5370	7820	3585	4540	4755	5225	4530	10	5	0	5	0	22	0	0	0	0	8
RRD-13	3980	3295	5250	1400	1770	3515	1440	930	920	1045	635	89	0	82	20	0	35	33	0	26	15	0
RRD-14	4960	3745	6345	1965	1710	4525	1870	1385	1210	1695	850	123	330	162	40	0	345	220	0	20	0	0
RRD-15	7530	8780	5740	9660	5850	9040	5160	4540	4070	5030	4100	0	0	0	0	0	0	0	0	0	0	0
RRD-16	10170	11600	4355	7805	4770	7770	3992	3150	2495	3650	2420	221	356	230	155	0	431	0	0	189	224	52
RRD-17	4880	3850	6090	2430	2285	5210	1795	1255	1190	2040	925	122	0	114	0	0	123	0	15	0	0	76
RRD-18	3160	2025	3850	1175	1330	2340	640	425	210	580	305	65	22	221	267	188	0	6	0	3	7	21
RRD-19	3870	3200	4640	1350	1790	3905	1420	1095	965	1475	640	0	0	0	0	0	0	0	0	0	0	0
RRD-20	2900	2300	3110	1450	1250	2270	950	630	530	590	430	0	0	0	0	0	0	0	0	0	0	0

Figure 3. Construct Data

2.3.4 Data Integration

At this stage is the data that will be used in stages of modeling is data transaction in may 2016 which have undergone the process attribute changes, here are some of the attributes are changed:

a. sweet bread 1000 becomes Roti1
b. sweet bread 2000 become Roti2
c. Cake Pia became Pia
d. the remaining 1000 be Sroti1 sweet bread
e. rest of sweet bread 2000 become Sroti2
f. the remainder of Cake Pia became Spia

Then the results of these changes combined with the attributes of the contents of the variant are connected by the mark of an underscore between the two attributes. After that is done the merger of two or more tables that have different information about the same object into a new set of data.

2.4 Modelling

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 Modelling Technique

Modeling techniques that will be used i.e. using hierarchical clustering method. The model that will be used was Hierarchical Agglomerative Clustering algorithm (AHC).

2.4.2 Mechanisme of Modelling

At this point mejelaskan concerning the mechanism of the model that will be used as well as the initial assumptions of the model. Agglomerative hierarchical clustering algorithm model has the following stages:

1. any data objects are considered as a group. If n = number of data and k = the number of groups, then k = n.
2. Calculate the distance matrix between groups (using the euclidean distance).
3. Locate the two groups who have the closest distance (complete linkage) then combine the two groups so that k = k-1.
4. If k > 1, repeat steps 2 and 3.
5. If k = 1, the iteration is complete.
6. Generate group by cutting the dendrogram at the right level.

2.4.3 Using of Modelling Technique

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

Change any data objects into groups

Initialization initial group is as much as 20 groups, due to data (n) = Group (k), then calculated the distance between the group with existing groups by using Euclidean Distance equations.

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

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 size:

$$d(RRD-01, RRD-20) = \sqrt{\begin{array}{c}(7490-2900)^2+(5830-2300)^2+(4590-3110)^2+\\(6502-1450)^2+(4215-1250)^2+(6400-2270)^2+\\(2640-950)^2+(3320-630)^2+(3460-530)^2\\+(4170-590)^2+(3230-430)^2+(0-0)^2+\\(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2\\+(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2\\+(0-0)^2\end{array}} = 11243,83$$

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.

2nd step

In this stage, each group will be calculated distance with himself and with other groups. The following is the process of the calculation of the distance between groups by using Euclidean Distance equations. The following is an example of the calculation of the distance by total:

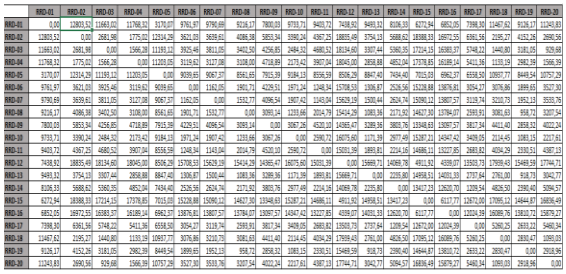


Figure 4. *Euclidean Distance Matrix*

Step 3

Look for the two groups to each other.

At this stage will be calculated the distance between the two groups by using complete linkage. By using this method, it will look for the closest distance from the two groups. The following is an example of two groups to each other.

Following is the result of some calculation iteration is done.

Iteration 1: d _ (13,19) 1 = max {8537.64, 8363.57} = 8537.64

Up to:

d _ (u, v) n = max {d _ ((u, n)), d _ ((v, n))}

.... Equation 3.1

With complete linkage method on setap iteration two groups which has a very small distance will be merged. If the Group of (n, n) and euclidean distance = 0, then the value of the euclidean distance is not declared as the smallest and smallest distance search will continue.

At the time of the grouping, the first iteration of the closest distance is 13 and 19 with the value to the distance i.e. 918.72. Then the group will be combined into one group and the number of new groups formed into as many as 19 of the group. The following 19 groups and euclidean distance for each group.

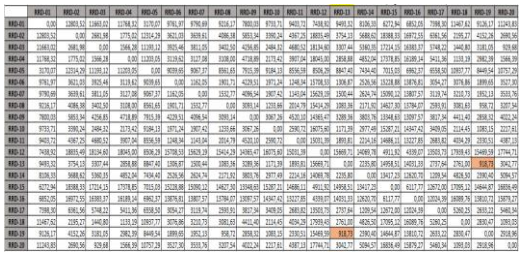


Figure 5. Euclidean Distance matrix in first iteration

The 4th step

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.

5th step

When the number of clusters amounted to 1, then the phase 3 has been completed and the final result of iteration based on the following transactions is a dendogram of the grouping results. as shown in Figure 6. Final Results Of The Dendogram

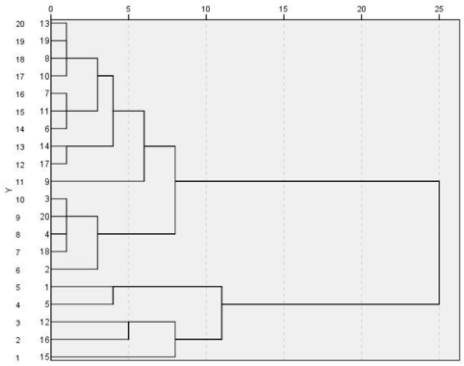


Figure 6. Final Result of Dendogram

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 cluster is formed. To know the best number of clusters, one of which with a use agglomeration schedule. Agglomeration schedule can be seen in Figure 7.

Iterasi	Kombinasi Kelompok		Koefisien	Iterasi Kombinasi Kelompok sebelumnya		Iterasi Selanjutnya
	Kelompok 1	Kelompok 2		Kelompok 1	Kelompok 2	
1	RRD-13	RRD-19	918,727	0	0	3
2	RRD-03	RRD-20	929,677	0	0	9
3	RRD-08	RRD-13	1083,150	0	1	7
4	RRD-4	RRD-18	1133,188	3	0	9
5	RRD-07	RRD-11	1143,042	2	0	8
6	RRD-14	RRD-17	1209,541	5	0	13
7	RRD-08	RRD-10	1233,660	0	0	10
8	RRD-06	RRD-07	1248,344	0	7	10
9	RRD-03	RRD-04	1566,393	0	0	11
10	RRD-06	RRD-08	2590,716	8	4	13
11	RRD-02	RRD-03	2690,558	0	6	17
12	RRD-01	RRD-05	3170,066	11	10	18
13	RRD-06	RRD-14	3409,053	12	9	15
14	RRD-12	RRD-16	4339,074	0	0	16
15	RRD-06	RRD-09	4520,096	13	0	17
16	RRD-12	RRD-15	6117,774	0	0	18
17	RRD-02	RRD-06	6361,557	16	0	19
18	RRD-01	RRD-12	8506, 293	14	15	19
19	RRD-01	RRD-02	18835,490	18	17	0

Figure 7. *Agglomeration Schedule*

The number of clusters the best can be seen from the difference in coefficient of each iterasinya. The difference of the segmentation based on coefficient sizes can be seen on Figure 8.

Jumlah Kelompok	Terakhir *Agglome*	Koefisien Tahap ini	Selisih
2	18835,490	8506,293	10329,197
3	8506,293	6361,557	2144,736
4	6361,557	6117,774	243,783
5	6117,774	4520,096	1597,678
6	4520,096	4339,074	181,022
7	4339,074	3409,053	930,021
8	3409,053	3170,066	238,987
9	3170,066	2690,558	479,508
10	2690,558	2590,716	99,842
11	2590,716	1566,393	1024,323
12	1566,393	1248,344	318,049
13	1248,344	1233,660	14,684
14	1233,660	1209,541	24,119
15	1209,541	1143,042	66,499
16	1143,042	1133,188	9,854
17	1133,188	1083,150	50,038
18	1083,150	929,677	153,473
19	929,677	918,727	10,950

Figure 8. 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 the Figure 8

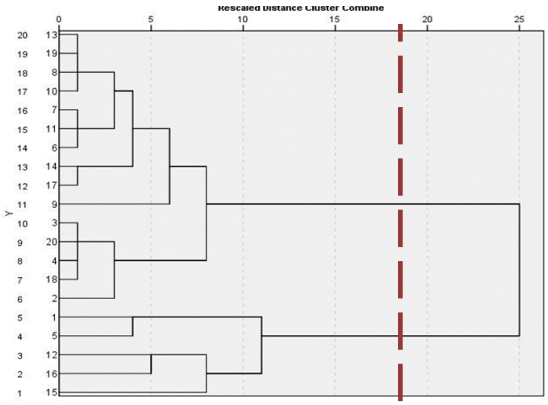


Figure 9. Cutting of Dendogram

The end result of this grouping, formed two groups with each group can be seen in table 3 and table 4.

Table 2. *Cluster* 1

ID	Nama of owner	Address
RRD-02	Asep Rachman	Ranca Ekek
RRD-03	Aris Winarso	Bale Endah
RRD-04	Goyon Winarno	Cicalengka
RRD-06	Marno Suroto	Majalaya
RRD-07	Ateng Sanjaya	Cibiru
RRD-08	Tursino Widagdo	Majalaya
RRD-09	Imas Darmaji	Majalaya
RRD-10	Ferdi Septian	Garut
RRD-11	Saprol Mahmud	Majalaya
RRD-13	Gilang Rambu	Garut
RRD-14	Suyoto	Sumedang
RRD-17	Yono Fathah	Sumedang
RRD-18	Lusyuno Hadi Kusumo	Ranca Ekek
RRD-19	Otong Sirait	Sumedang
RRD-20	Tino Wijaya	Cilacap

Table 3. *Cluster* 2

ID	Nama of owner	Address
RRD-01	Toni Atmajaya	Purwakarta
RRD-05	Abas Yahya	Majalaya
RRD-12	Toyo Faturakhman	Tasik
RRD-15	Atam Gunardi	Karawang
RRD-16	Wasko Satrio	Lembang

From the results of the grouping of distributors, obtained the lowest refund

Tabel 4. Pengembalian Terendah

Kelompok	Jenis Produk
Cluster 1	Sroti2_Meses
	Sroti2_Kombinasi
	Sroti2_Coklat
Cluster 2	Spia_Coklat
	Sroti1_Kopyor
	Sroti2_Meses

Results from the lowest returns, product recommendations can be generated for each cluster can be seen in table 5 below:

Table 5. Recomendation of Products

Cluster	Rekomendation
Cluster 1	Give priority to selling Roti 2000 Meses, Roti 2000 Kombinasi, Roti 2000 Coklat.
Cluster 2	Give priority to selling Kue Pia Coklat, Roti 1000 Kopyor, Roti 2000 Meses.

3. CONCLUSION

Results from research that has been done, the conclusion that the application is a grouping of distributors who have built a marketing manager can help to form groups that correspond to the transaction so that product likeness referable to distribution and business strategies appropriate to the characteristics of each of its distributors.

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

1. Creation of new variant performed Bread Riki in quite a long time. Therefore, to anticipate the need of adding the attribute settings feature, so that if there is a new variant, the formation of groups can be disesuai with the addition of a new variant.
2. the method of grouping can be developed again using another algorithm or combine 2 different algorithms, such as combining algorithm AHC (Agglomerative Hierarchical Clustering) and K-Means algorithm, so that it can be compared where a better algorithm in the grouping.
3. Methods of AHC used can be expanded by using the Average linkage, so that it can be compared to other methods of mengan which more accurately to calculate the distance.
4. The methods used in the calculation of distances can be developed again by using the method of Manhattan, so it can be compared to other methods of calculation.

REFERENCES

[1] D. S. Rajagopal, "Customer Data Clustering Using Data Mining Technique," International Journal of Database Management Systems (IJDMS), vol. 3, no. 4, pp. 1-3, 2011.

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

[3] D. T. Larose, Discovering knowledge in data, New Jersey: John Wiley & Sons, 2005.

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

[5] R. W. P. Chapman, 2000. [Online]. Available: http://the- modeling-agency.com/crisp-dm.pdf. [Diakses 23 Agustus 2016].

[6] Fathansyah, Basis Data, Bandung: Informatika, 2015.

[7] K. Tampubolon, H. Saragih dan B. Reza, "Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-alat Kesehatan," Majalah Ilmiah Informasi dan Teknologi Ilmiah (INTI), vol. 1, no. 1, 2013.

[8] L. (Allium ursinum L., "Research Gate," November 2015. [Online]. Available: https://www.researchgate.net/figure/284724792_fig2_Fig-2-A-Dendrogram-AHC-analysis-representing-dissimilarity-relationships-of-14. [Diakses 29 October 2016].

[9] I. Sommerville, Software Engineering (9th Edition), Pearson, 2010.

[10] I. Supardi, Semua Bisa Menjadi Programming Java Basic Programming, Jakarta: Elex Media, 2010.

[11] B. Hambling dan P. v. Goathem, User Acceptance Testing : Step by Step Guide, Swindon: BCS Institute, 2013.

[12] T. Alfina, B. Santosa dan A. R. Barakbah, "Analisa Perbandingan Metode Hierarchical," JURNAL TEKNIK ITS, vol. 1, p. 1, 2012.

[13] R. Ramakrishnan dan J. Gehrke, Database Management System, Singapore, 2003.