Application of Data Mining in Product Sales Data to Form Outlet Segmentation in Blackjack Industries Bandung

Faza Rahthin A[1], Dian Dharmayanti [2]
[1,2]Informatika - Universitas Komputer Indonesia
Jl. Dipatiukur 112-114 Bandung
E-mail : fazarahthin@gmail.com[1], diandmdr@yahoo.com[2]

ABSTRAK

Blackjack Industries is a manufacturer of apparel that includes t-shirts, shirts and bags to market its products at outlets scattered in some areas of Indonesia. Blackjack Industries new product submit every once a month to every outlet and outlet accept the product and then market to customers. The new product in question is a category of product types and models. Based on the questionnaire survey that has been conducted against the 20 outlets Blackjack Industries note that 14 outlet experience constraints in terms of product category and type mismatch model which means that product category types and models that are sold at each outlet does not correspond to the needs of the customer. This led to many categories of product types and models that are not sold. Then type the product category and model required each outlet is different and requires just the right characteristics for each outlet.

The availability of data, the need for an information (or knowledge) that are useful as supporting decision making. The data obtained are useful in getting the exact characteristics of every outlet. Characteristics in the can by doing an analysis of the data product sales to each outlet with the use of Data Mining. One of the methods contained in data mining are used in this research is clustering. where the technique used Agglomerative hierarchical clustering algorithm using (AHC) in the establishment of appropriate characteristics in the segmentation of sales data.

Based on the results of the testing that has been done can be inferred that this outlet segmentation can help Parties Blackjak Industries in determining the needs of each outlet based on characteristics that are in can and ease the marketting manager in determining the right product recommendations.

Keywords: Hierarchical Agglomerative Clustering, Data Mining, Clustering, segmentation, Outlet

INTRODUCTION

1.1 Background

Blackjack Industries is a manufacturer of apparel that includes t-shirts, shirts, bags to market its products at outlets scattered in some areas of Indonesia. The main task of the outlet of Blackjack that is serving the outlets directly, so the existence of this outlet is very important to be able to define the business strategy Blackjack Industries in the future.

After conducting interviews with the marketting Blackjack Industries, obtained information that Blackjack Industries new product send out every once a month to every outlet and outlet accept the product and then market it to the outlet. All outlets will receive the same amount of product and each delivery is made. Based on the questionnaire survey that has been conducted against the 20 outlets Blackjack Industries note that 14 outlet experience constraints in terms of product discrepancies. Is the mismatch of the product is a product that is in the outlet does not comply with the required outlet such as a lack of interest in selling products that each outlet that judging from the type and model of the resulting abundance of products that are not purchased by the outlet. This causes a lot of products that are not sold. The conclusion in May that products needed for each different outlets.

Based on an issue that has been presented before, a possible solution is performed by forming segmentation outlet so that the characteristic needs of each outlet will be visible. Needs of each outlet can be seen based on sales data which have been made at each outlet. But only by looking at the sales data in the sense of less clear in looking at the characteristics of the products of each outlet. In order to clarify the characteristics of the products of each outlet required sales data, data products and data outlets. With a view of the data can be a marketting for the knowledge in analyzing different characteristics or behavior that may require a different product or marketing of any segment of the outlet. So are expected to maximize the process of marketing the right product to each outlet.

The technique of clustering can be used to mensegmentasikan outlet based on proximity of characteristics between outlet [1]. One method that can be used in the technique of clustering Agglomerative Hierarchical Clustering is [8]. Such methods will be used in this research because with reference from an existing reference method will perform the grouping against the cluster or objects

that are not yet known. In addition, cluster outlet is formed naturally based on the characteristics of the product.

Therefore consideration and explanation of the above describe researchers interested in conducting research in Blackjack Industries.

Based on the discussion above, the intent of this research is to apply data mining on data sales, product data and data products to form the outlet outlet segmentation in Blackjack Industries.

While the goals to be achieved for this research is:

1. To assist the parties in determining the Industries Blackjak needs of each outlet based on characteristics that are in the can.
2. Facilitate the marketting manager in determining the right product recommendations.

2 THEORETICAL BASIS

2.1 Foundation of Theory

The cornerstone of the theory on writing this thesis will explain about the theories related to application segmentation outlet Blackjack Industries.

2.1.1 Data

The data is a representation of a real-world fact that represents an object such as a human (employees, students, shoppers, customers), goods, animals, events, concepts, conditions, etc., are recorded in the form of numbers, letters, symbols, text, image, sound, or a combination [5].

2.1.2 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 dredging, data archaeology [5].

2.1.3 CRISP-DM

CRISP – DM (Cross – Industry Standard of Data Mining). CRISP – DM data mining methodology is described with a model of the process and in it there are tasks that must be performed to do data mining project [2]

2.1.4 Clustering

Clustering is the work of separating data/vector into a number of groups (clusters) according to its characteristics respectively. Data – data that has the resemblance of a characteristic that converged in the same cluster, and data – data with different characteristics will be separated in different clusters [4].

Calculate the distance by using Euclidean:

$$d = \sqrt{\sum_{i=1}^{r}(X\,i - Yi)^2}$$ Equation 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

2.1.5 Agglomerative Hierarchical Clustering (AHC)

Agglomerative hierarchical clustering algorithm is used to measure the closeness between the two clusters. There are three ways to use AHC, including, single linkage method, complete linkage, average linkage. [9].

Here is an equation the maximum linkage:

(,)=maks{ (,)}; () ∈ equations 2

Here is an Agglomerative algorithm of stages:

1. Represent each outlet as one group, then calculate the distance (distance measure) between the groups.
2. Locate the two clusters that have the closest distance and select the greatest value to determine the new value of the two clusters that formed using complete linkage found in equation (2).
3. Repeat step two to left one group (a collection of all outlets) will stop 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 and the final result of looping based on size can be seen in attachment. The following is the dendrogram of the results of the grouping

2.1.6 Agglomeration Schedule

Agglomeration schedule is a numerical summary of cluster solutions resulting from the clustering process using hierarchical methods [12].

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. combined Cluster is a cluster which is joined on each stagenya.
3. Cooficients 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.

2.2 analysis of Problems

Based on the results of the interviews have been conducted with the marketing then obtained the following results:

1. There is an obstacle in the event of discrepancies, the product of the sale of products to customers.
2. There is No determination of an appropriate product recommendations.

Based on the analysis of the problem that has presented methods that are appropriate to the resolution of the case by using the method of Data mining. In this case it will use the process model

CRISP-DM. Stages CRISP-DM process model as follows:

1. Understanding the business process

The first stage in the process model of CRISP-DM is a stage of understanding the business (business understanding). At this stage, the collection of information concerning the business objectives and the expected results by marketing manager Blackjack Industries. Activities at this stage include identification of business goals, the determination of the purpose of Data mining.

Business goals of Blackjack Industries is increasing product quality at each Outlet. It is done to find out the characteristics for each type of products, especially products that match the needs of customers in each Outlet. By doing this it is hoped customers would purchase back and buy more products.

Business success criteria according to the Manager of marketing, important aspects that became criteria to achieve the business objectives above is by selling the right product in accordance with the characteristics of each Outlet to customers, to do so then the Blackjack Industries need to do an analysis to every Outlet and determine the appropriate products for each Outlet. The analysis was conducted using Data products, and sales Outlet so that the parties can determine the marketting business strategy that's right for the future.

1.1.1 Pinpointing Data Mining

This activity aims to describe the expected results of the research, so that business objectives can be achieved. At this stage the described project goals from the point of view of engineering. Following is the result from the determination of target Data mining.

a. the purpose of Data mining

The purpose of the application of Data mining on the sale of products in the research is to make up the segmentation Data Outlet based on sales of products that have been conducted by the Outlet. With the application of Data mining is expected to assist the parties in determining the marketing strategy of the company's business.

b. criteria for successful Data mining

The criteria of success against this research was able to help the parties grouped in the Blackjack Industries marketing every Outlet has the same characteristics in the product sales.

2.2 Understanding the Data

Stages of understanding Data is the second stage after stage of understanding the business. In the stages of understanding this Data there are a number of steps including

2.2.2 Explanation of Data

The data results of the join has a whole 10 attributes, 2 attributes about the Outlet, 5 types of explanation regarding the attributes and motives of the products sold.

2. the preparation of the initial Data 2.3.

On initial Data preparation process will be prepared Data to be used in the modeling phase. The data will be used in stages of modeling is Data the results of the join that have undergone selection process attributes. The data attribute of the election results is a Data set that can be seen in table 1.

Table 1. The Preparation Of The Initial Data

ID Outlet	TS_OBLG	TS_RGLN	KMJ_PLS	KMJ_FLN	JKT_HOD	JKT_PRK	SWT_CRDG	SWT_RJT	POLO_PLS	BAG_PLS
1	9	8	4	11	4	5	4	2	8	11
2	13	11	8	11	7	4	5	4	10	7
3	7	13	2	15	5	7	3	5	9	8
4	15	8	7	12	5	5	3	4	8	6
5	14	12	3	8	3	6	5	1	10	8
6	9	11	5	9	6	8	4	4	9	3
7	14	13	8	9	5	3	7	7	11	13
8	11	7	3	14	6	6	3	6	9	8
9	10	13	5	11	4	6	5	2	9	9
10	13	11	8	13	5	6	5	5	9	9
11	13	14	6	13	6	5	5	5	9	11
12	10	13	7	12	10	4	6	3	9	9
13	14	11	5	12	9	7	4	3	8	7
14	14	8	6	8	1	8	3	5	8	7
15	10	10	6	8	7	3	4	5	11	9
16	11	12	4	13	6	5	6	3	11	9
17	7	9	5	11	4	6	5	2	9	9
18	13	13	8	11	7	4	5	4	10	7
19	11	9	2	15	5	7	3	5	9	8
20	11	8	3	14	6	6	3	6	9	8

Data Set Description

On the description of the Data will be in the explain the naming attribute. For an explanation of the names of the attributes in the Data set will be in explain below:

Description the table name Attribute from the Data Selection Results the results of the Join (2):

TS = type of T-shirts
KMJ = type of shirt
JKT = type of Jacket
SWT = type of Sweater
POLO = polo shirt Type
BAG = bags
OBLG = Model T-shirts
RGLN = Model Raglan
PLS = Plain Models
FLN = Model Flanner
HOD = Models Hoodie
PRK = the Model Parka
CRD = Cardigan Model
RJT = Model Raju

2.4 Stage Modeling

The next stages of the framework of CRISP-DM is the stages of modeling, in this stage select and prepare the appropriate modeling with the aim of the treetops.

2.4.1 the selection method of Modeling

The first step in the modeling is to choose a method that will be used. The selection methods tailored to the Data has been prepared. The method to be used is a method of hierarchical agglomerative clustering.

After modeling method is specified, then it will be explained the procedure of such methods. The steps of the method of Agglomerative Hierarchical clustering (AHC) is as follows.

1. Represent each Outlet as a segmentation, then calculate the distance (distance measure) between segmentations.
2. Locate the two clusters that have the closest distance and select the greatest value to determine the new value of the two clusters that formed using complete linkage found in equation 2.
3. Repeat step two until one remaining segmentation (collection of all Outlets) will stop 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 and the final result of looping based on size can be seen in attachment.
5. Generate a dendrogram with clusters of chop on the right level. A dendrogram based on already established, the right level or number of clusters would best be determined by looking at the difference between the distance from each cluster are formed. To find out the number of clusters the best way one can use agglomeration schedule.

Step-1

Initialization initial cluster is as much as 20 cluster because the Data (n) = cluster (k), then calculated the distance between the cluster with an existing cluster using the Euclidean Distance equation (Equation 1). each cluster will be calculated distance with himself and the gap with other clusters. This is a process of calculation of distances between clusters by using Euclidean Distance equation (Equation 1). The following is an example of the calculation of the distance by size:

Dman (1, 2) == 8.72

In the same way i.e. by using Euclidean Distance equation calculation is performed on the entire Data. Process calculation of Euclidean Distance matrix is generated then it will distance.

2nd step

Search two clusters that have the closest distance

At this stage will be the distance between two clusters is computed using complete linkage (equation 2), with this method will look for the closest spacing of two clusters and cluster merging with maximum distance and generate new cluster.

Here's an example:

Table 2 Euclidean Distance Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	8,72	8,60	8,77	8,25	9,80	11,05	6,86	5,83	7,87	8,66	9,11	9,06	9,06	7,28	6,78	3,61	9,59	7,07	6,78
2	8,72	0,00	10,58	5,39	8,12	8,00	8,00	8,31	6,63	4,24	6,08	5,57	5,48	9,17	5,57	5,83	8,43	2,00	8,94	7,87
3	8,60	10,58	0	11,53	11,22	9,06	13,64	7,55	7,07	9,27	8,66	9,54	9,70	12,57	10,44	6,93	7,55	10,39	7,81	7,81
4	8,77	5,39	11,53	0	8,66	8,89	11,09	6,93	8,66	5,57	8,60	9,80	6,08	6,71	8,60	9,80	6,08	6,71	8,60	8,43
5	8,25	8,12	11,22	8,66	0	8,94	10,30	10,44	5,83	8,60	8,66	10,44	8,25	7,62	8,54	7,21	8,66	8,12	9,70	9,49
6	9,80	8,00	9,06	8,89	8,94	0	13,64	9,11	7,62	9,17	10,82	9,54	7,87	8,94	8,43	8,72	7,81	8,25	9,06	8,72
7	11,05	8,00	13,64	11,09	10,30	13,64	0	12,29	9,38	7,62	6,56	9,43	11,14	11,66	7,94	8,94	11,70	7,75	12,73	11,83
8	6,86	8,31	7,55	6,93	10,44	9,11	12,29	0	8,66	7,28	8,83	9,80	7,42	9,33	8,49	7,14	7,62	9,43	7,81	7,28
9	5,83	6,63	7,07	8,66	5,83	7,62	9,38	8,66	0	6,00	5,74	6,86	7,35	9,06	7,14	4,24	5,00	6,32	7,62	6,32
10	7,87	4,24	9,27	5,57	8,60	9,17	7,62	7,28	6,00	0	4,36	7,00	6,16	8,25	7,55	5,66	7,94	4,69	7,35	7,28
11	8,66	6,08	8,66	8,60	8,66	10,82	6,56	8,83	5,74	4,36	0	6,16	6,86	10,82	8,00	5,00	9,17	5,39	8,19	8,06
12	9,11	5,57	9,54	9,80	10,44	9,54	9,43	9,80	6,86	7,00	6,16	0	6,56	13,15	6,93	5,74	8,49	5,20	9,95	9,22
13	9,06	5,48	9,70	6,08	8,25	7,87	11,14	7,42	7,35	6,16	6,86	6,56	0	9,75	10,68	9,54	10,00	9,70	9,27	8,83
14	9,06	9,17	12,57	6,71	7,62	8,94	11,66	9,33	9,06	8,25	10,82	13,15	9,75	0	7,21	6,24	9,64	8,19	6,40	8,19
15	7,28	5,57	10,44	8,60	8,54	8,43	7,94	8,49	7,14	7,55	8,00	6,93	10,68	7,21	0	6,86	7,21	6,24	9,64	8,19
16	6,78	5,83	6,93	8,43	7,21	8,72	8,94	7,14	4,24	5,66	5,00	5,74	9,54	6,24	6,86	0	6,40	5,83	6,32	5,83
17	3,61	8,43	7,55	9,49	8,66	7,81	11,70	7,62	5,00	7,94	9,17	8,49	10,00	9,64	7,21	6,40	0	9,11	7,55	7,42
18	9,59	2,00	10,39	6,71	8,12	8,25	7,75	9,43	6,32	4,69	5,39	5,20	9,70	8,19	6,24	5,83	9,11	0	9,59	8,94
19	7,07	8,94	5,66	7,81	9,70	9,06	12,73	3,00	7,62	7,35	8,19	9,95	9,27	6,40	9,64	6,32	7,55	9,59	0	1,00
20	6,78	7,87	6,78	6,86	10,00	8,72	11,83	1,00	8,00	6,78	8,06	9,22	8,86	5,83	8,19	6,48	7,42	8,83	1,00	0

(dman) = (d (8.20)) = 1. Of all the existing clusters will be chosen two different cluster as cluster 1 and 2 by looking at the very minimum of the value of all of the clusters. Two of the same cluster as cluster 1 and 1 or 2 and 2 not selected because it is the same cluster and has a minimum value that is 0. For the early election of the chosen clusters clusters 8 and 20 because it has the most value for a minimum of another cluster. Calculate the distance 8 and 20 with another cluster. Example as follows:

d (8.20) 4 = Max {d (8.4) d (22.3)} = max {(6,93) (6,86)} = 6.93 because after comparing the 2 values i.e. 6.93 6.86 and selected the most maximum value. After doing calculations with other segmentation Cluster d (8.20) will form a new cluster and cluster 8 and 20 will be delete. And will appear a new cluster i.e. 8.20. in the election the most value this maximum is done to get a new value to the new cluster. Example as follows:

Table 3 Euclidean Distance Matrix Iteration-1

	1	2	3	4	5	6	7	8,20	9	10	11	12	13	14	15	16	17	18	19
1	0	8,72	8,60	8,77	8,25	9,80	11,05	6,86	5,83	7,87	8,66	9,11	9,06	9,06	7,28	6,78	3,61	9,59	7,07
2	8,72	0,00	10,58	5,39	8,12	8,00	8,00	8,31	6,63	4,24	6,08	5,57	5,48	9,17	5,57	5,83	8,43	2,00	8,94
3	8,60	10,58	0	11,53	11,22	9,06	13,64	7,55	7,07	9,27	8,66	9,54	9,70	12,57	10,44	6,93	7,55	10,39	5,66
4	8,77	5,39	11,53	0	8,66	8,89	11,09	6,93	8,66	5,57	8,60	9,80	6,08	6,71	8,60	9,80	6,08	6,71	7,81
5	8,25	8,12	11,22	8,66	0	8,94	10,30	10,44	5,83	8,60	8,66	10,44	8,25	7,62	8,54	7,21	8,66	8,12	9,70
6	9,80	8,00	9,06	8,89	8,94	0	13,64	9,11	7,62	9,17	10,82	9,54	7,87	8,94	8,43	8,72	7,81	8,25	9,06
7	11,05	8,00	13,64	11,09	10,30	13,64	0	12,29	9,38	7,62	6,56	9,43	11,14	11,66	7,94	8,94	11,70	7,75	12,73
8,20	6,86	8,31	7,55	6,93	10,44	9,11	12,29	0	8,66	7,28	8,83	9,80	7,42	9,33	8,49	7,14	7,62	9,43	7,81
9	5,83	6,63	7,07	8,66	5,83	7,62	9,38	8,66	0	6,00	5,74	6,86	7,35	9,06	7,14	4,24	5,00	6,32	7,62
10	7,87	4,24	9,27	5,57	8,60	9,17	7,62	7,28	6,00	0	4,36	7,00	6,16	8,25	7,55	5,66	7,94	4,69	7,35
11	8,66	6,08	8,66	8,60	8,66	10,82	6,56	8,83	5,74	4,36	0	6,16	6,86	10,82	8,00	5,00	9,17	5,39	8,19
12	9,11	5,57	9,54	9,80	10,44	9,54	9,43	9,80	6,86	7,00	6,16	0	6,56	13,15	6,93	5,74	8,49	5,20	9,95
13	9,06	5,48	9,70	6,08	8,25	7,87	11,14	7,42	7,35	6,16	6,86	6,56	0	9,75	10,68	9,54	10,00	9,70	9,27
14	9,06	9,17	12,57	6,71	7,62	8,94	11,66	9,33	9,06	8,25	10,82	13,15	9,75	0	7,21	6,24	9,64	8,19	6,40
15	7,28	5,57	10,44	8,60	8,54	8,43	7,94	8,49	7,14	7,55	8,00	6,93	10,68	7,21	0	6,86	7,21	6,24	9,64
16	6,78	5,83	6,93	8,43	7,21	8,72	8,94	7,14	4,24	5,66	5,00	5,74	9,54	6,24	6,86	0	6,40	5,83	6,32
17	3,61	8,43	7,55	9,49	8,66	7,81	11,70	7,62	5,00	7,94	9,17	8,49	10,00	9,64	7,21	6,40	0	9,11	7,55
18	9,59	2,00	10,39	6,71	8,12	8,25	7,75	9,43	6,32	4,69	5,39	5,20	9,70	8,19	6,24	5,83	9,11	0	9,59
19	7,07	8,94	5,66	7,81	9,70	9,06	12,73	3,00	7,62	7,35	8,19	9,95	9,27	6,40	9,64	6,32	7,55	9,59	0

3rd step

Repeat step two until one remaining segmentation (collection of all Outlets) step will stop when the number of clusters are formed is equal to 1.

The 4th step

When the number of clusters amounted to 1 then the 3rd stage has been completed and the final result of looping based on size can be seen in attachment. The following is the dendrogram of the segmentation results. Dendgoram results in Figure 2 Dendogram end results as follows:

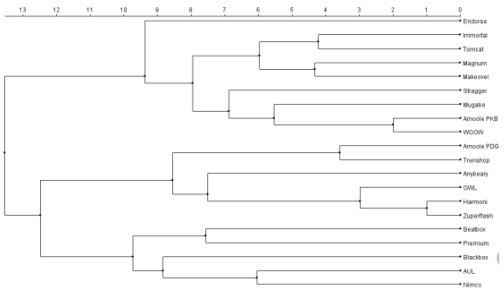


Figure 3.1 the end result Dendogram

5th step

Generates cluster by cutting the right level on the dendrogram

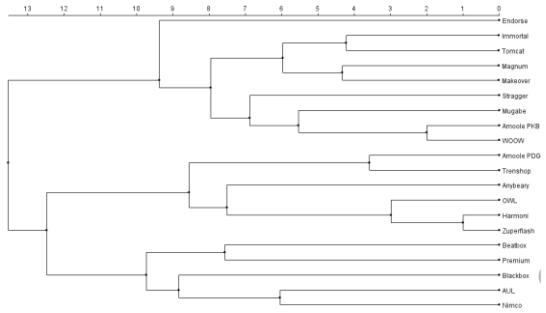
A dendrogram based on already established, the right level or number of clusters would best be determined by looking at the difference between the distance from each cluster are formed. To find out the number of clusters the best way one can use agglomeration schedule. Agglomeration schedule can be seen in Table 3.12.

The number of clusters the best can be seen from the difference koofisien from any iterasinya. The difference of koofisien can be seen in Table 4.

Table 4 Difference Koofisien Segmentation by size

Jumlah Cluster	Tahap Terakhir Agglomeration	Koefisien Tahap ini	Selisih
2	13,64	12,57	1,07
3	12,57	9,75	2,82
4	9,75	9,43	0,32
5	9,43	8,89	0,55
6	8,89	8,60	0,29
7	8,60	8,00	0,60
8	8,00	7,62	0,38
9	7,62	7,55	0,07
10	7,55	6,93	0,62
11	6,93	6,08	0,85
12	6,08	6,00	0,08
13	6,00	5,57	0,43
14	5,57	4,36	1,21
15	4,36	4,24	0,12
16	4,24	3,61	0,64
17	3,61	3,00	0,61
18	3,00	2	1,00
19	2	1	1

The biggest difference that is bringing the total cluster size was 2.82 is best 3 clusters. Then, the dendrogram based on segmentation result size will be truncated as shown in Figure 5



Tabel 4 *Agglomerative Schedule*

Iterasi	Kombinasi Cluster		Koefisien	Iterasi Kombinasi Cluster sebelumnya		Iterasi Selanjutnya
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	8	20	1	0	0	3
2	2	18	2,00	0	0	7
3	8	19	3,00	1	0	11
4	1	17	3,61	0	0	14
5	9	16	4,24	0	0	8
6	10	11	4,36	0	0	8
7	2	12	5,57	2	0	10
8	9	10	6,00	5	6	13
9	4	13	6,08	0	0	15
10	2	15	6,93	7	0	13
11	3	8	7,55	0	3	14
12	5	14	7,62	0	0	17
13	2	9	8,00	10	8	16
14	1	3	8,60	4	11	18
15	4	6	8,89	9	0	17
16	2	7	9,43	13	0	19
17	4	5	9,75	15	12	18
18	1	4	12,57	14	17	19
19	1	2	13,64	18	16	0

From the results of the segmentation based on product sales Outlets obtained the highest results of any cluster that is used as a reference for the determination of the characteristics of each cluster, as follows.

The table was 3.20 the highest Sales on each Cluster

Cluster	**Jenis dan Model**	**Jumlah**
1	KMJ_FLN	15
2	TS-OBLG	14
	TS-RGLN	14
3	TS-OBLG	15

2.5 Evaluation

At this stage, the model is already formed and are expected to have good quality when viewed from the perspective of Data analysis. At this stage will be evaluated against the effectiveness and quality of the model before use and determine whether the model can be mencapat goals set at the initial phase (Business Understanding).

2.5.1 Results Evaluation

This measure assesses the extent to which models meet business goals and trying to determine whether there is some business reasons why this model lacked.

2.5.2 Assessment Data mining results with respect to the criteria of business success

The assessment is done by linking business success criteria with a model that is already formed. The criteria of the expected business success is to sell the right products in accordance with karaketeristik of every Outlet. While the models that already formed can be seen on the results of product recommendations.

2.5.3 Model approved

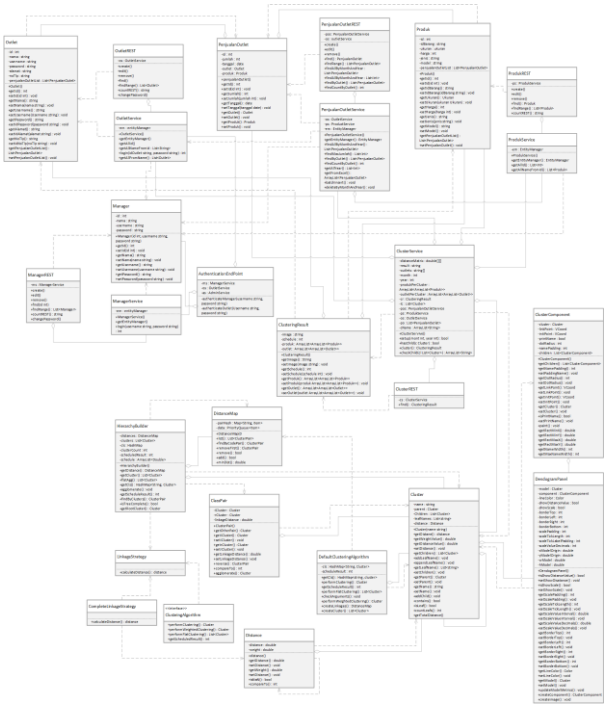
The approved model can meet the criteria for the success of the business. With a view of assessment data mining terhapa business success criteria then the conclusions is a model that's been formed that has the characteristics already meets the criteria of success of the business. By looking at the results of this model are then formed can be approved and can be used as a reference for the marketting in the determination of the future strategy.

2.6 Use Case Diagram

Use Case Diagrams are used to describe the functional requirements of the system to be built. Use Case Diagram describes the relationships between the user with the system. This diagram became the starting point for the development of systems using object-oriented approach. Figure 3.3 is a Use Case Diagram of segmentation system Outlet on this research.

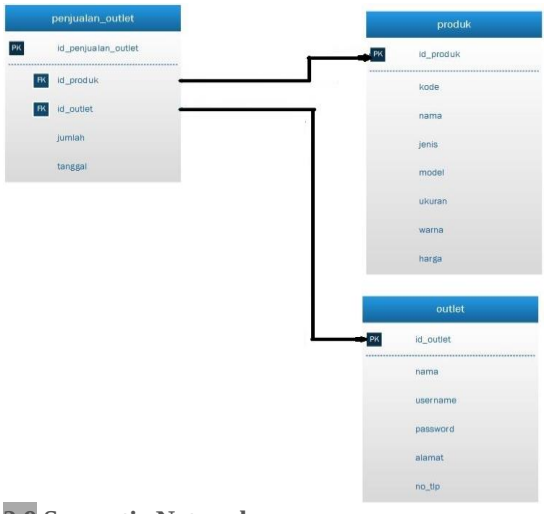
2.7 Class Diagram

Charts or diagrams describing the class structure of the system in terms of the definition of the classes that will be created to build the system. Classes have what are called attributes and methods or operations. Class diagrams are needed to build the system.



2.8 the scheme of Relations

The partnership scheme is a series of relationships between two or more tables in a database system. Here is a schematic relation on the system establishment of segmentation Outlet on Blacjack Industries:



2.9 Semantic Network

Semantic Network describes the relationship between the components of an interface consists of form and message. Image semantic networks 6 of the segmentation system Outlet.

1. the Semantic Network Outlet

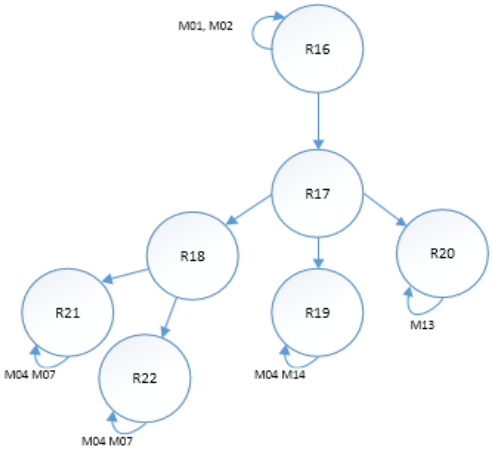


Image Semantic Network Outlet 3.56

1. the Semantic Network Marketting Manager

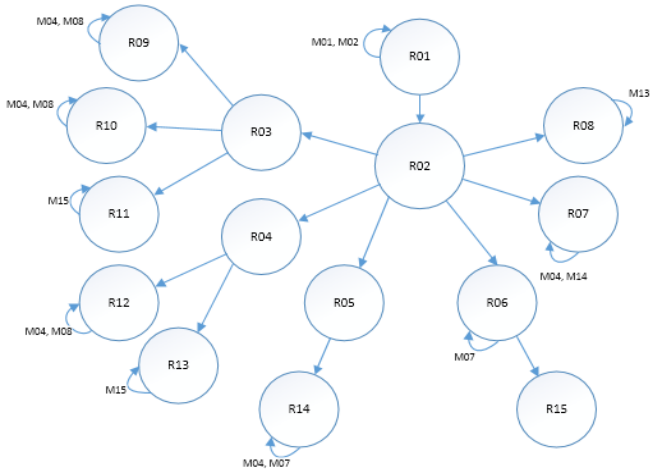


Figure 5 Semantic Network Marketting Manager

2.10 Test

Testing applications were made to the functionality of the application to assess whether the application meets the criteria already built or there is a discrepancy in functionality as that found on the design of the application. Testing is done to test the application is black box testing.

The testing plan will do is to how to test applications developed in black box and beta. Black box testing is focused on testing the functionality of the application is built.

2.10.1 Scenario Black Box

Black box testing is focused on the requirements of the functionality of the software to be built. Plan more testing can be seen in Table 7.

Table 7 the test Plan Application

Kelas Uji
2Login pengguna
Edit data produk
Hapus data produk
Tambah data produk
Hapus data outlet
Tambah data outlet
Edit data penjualan
Hapus data penjualan
Tambah data penjualan
Tambah data penjualan dengan excel
Lihat data penjualan
Melakukan *Clustering*

3. The COVER

From the results of this research can be concluded that

1. application the establishment of segmentation of the outlets have been built can be helpful in determining the Marketting Manager segmentation the outlets according to the characteristics of the product sales so it can be used as a reference for the right business strategy in the future.
2. Facilitate the Marketting Manager in determining the right product recommendations.

Pardew a few suggestions that may be made to the development of the application of the formation of these outlets, segmentation, among others:

1. The system may provide facilities for communication between the outlet with the admin manager marketing that is useful as a means of information.
2. In conducting the data analysis product sales manager marketing parties can choose the product category such as type, model, size and color to do the analysis, so that the system can determine the characteristics of the outlet of the selected product category.

BIBLIOGRAPHY

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

[2] C. Pete (NCR), et al., "*CRISP-DM 1.0 Step-by-step data mining guide*", SPSS, 2000.

[3] R.S. Arini dan M. Shalahuddin, "Modul Pembelajaran Rekayasa Perangkat Lunak (Terstruktur dan Berorientasi Objek)", Bandung: Modula, 2013.

[4] P. Eko., Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab, Yogyakarta: Andi, 2014.

[5] J. Han dan M. Kamber, *Data Mining* Concepts and Techniques, Jerman: Morgan Kauffman, 2006.

[6] A. Edgar, and C. Rodriguez. "The treatment of missing values and its effect on classifier accuracy." *Classification, clustering, and data mining applications*. Springer Berlin Heidelberg, 2004.

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

[8] B. Susanto, "*Data Preprocessin*g,", Versi 1.2, Jakarta: Elex Media Komputindo ,2013.

[9] P. Berkhin, *Survey of Clustering Data Mining Techniques*, Jerman: Accrue Software, 2006.

[10] Hermawati, Fajar Astuti, *Data Mining*, Yogyakarta: Andi, 2013.

[11] Xu, Rui, and Don Wunsch, *Clustering*, New Jersey: John Wiley & Sons. Inc, 2008.

[12] Caccam, Marielle., and Refran Jewel., *Clustering analysis, 2012, Wired*, 06 Desember 2012, [Online]. Tersedia: http://www.slideshare.net/jewelmrefran/*cluster*-analysis-15529464 [Diakses: 05 Januari 2017].

[13] L. Sofya, "Analisis *Cluster* dengan *Average Linkage Method* dan *Ward's Method* untuk Data Responden Nasabah Asuransi Jiwa Unit Link", pp. 84-91, Des 2014.