Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)
Edisi...Volume..., Bulan ..ISSN :

IMPLEMENTATION OF DATA MINING CLUSTERING METHOD IN BIOBSES BOOKSTORE USING K - MEANS ALGORITHM

Muhammad Alfariqi[1], Alif Finandhita[2]

[1,2] Teknik Informatika – Universitas Komputer Indonesia
Jl. Dipatiukur 112-114 Bandung

Email : alfariqimuhammad@gmail.com, alif.finadhita@yahoo.com

ABSTRAK

Biobses is a computer bookstore that has implemented e-commerce system on it transaction activities. The discounts granted system is also applied to e-commerce for each customer transactions. Unfortunately, the existing e-commerce is having difficulties on discount granted process especially on giving additional discounts to customers that is caused by ungrouping customer's data of their amount of transaction in each month.

Clustering method in data mining is the right solution to overcome the problem of grouping customer's data. By applying the k-means algorithm customer group formation begins by determining the number of clusters that will be formed, specifying the value of the centroid of each cluster, calculating the distance between the data and grouping the data by considering the minimum distance of the objects.

Through data mining applications with clustering method the grouping process of customer data on e-commerce system become easier. In addition, the granted of additional discounts problem can be overcome and it becomes easier with a discount system that can be customized in every month by company.

Keywords : Biobses, Book Store, Data Mining, Grouping, Transaction,

1. INTRODUCTION

Biobses is a company engaged in selling computer books from small to medium-sized markets throughout Indonesian. Bi-obses already using e-commerce system for marketing and sales of its products . e-commerce system is administered by the Division of Online with the main task of organizing online transactions , promotions and discounting.

Biobses book stores have two business process flows that is offline and online. In both of these business processes biobses stipulates a discount that has been limited by the company. The difference is in the flow of business transactions offline, when customers conduct transactions and spending minimal amount achieved by chance would be given an additional discount by the company. But in the online system which applied different companies, which the user will get an additional discount through their transactions for a month, if the shopping quota is reached, and granted an extra discount in the following month. But discounting is difficult to apply for online transactions in biobses that many per day. So get additional discounts granted to customers on their own can not be passed on to the end of the month. And the initial target marketing is finding many customers as possible who are assigned to the division line was not achieved in because of the difficulty to segment customers by their transactions.

One attempt to solve the problems of discounting is by grouping customers by using the principles of data mining. By doing data mining, companies can establish customer grouping bedasarkan transactions they have done and analyze the market to understand existing markets or find new opportunities to increase profits including predict marketing criteria. This technology now exists and is applied by companies to solve business problems (Achmad Zulfikar: 2009). Data mining is the mining or the discovery of new information by looking for patterns or specific rules on very large amounts of data (Davies 2004). One method is clustering data mining is the process of grouping data set into groups so that objects within a group have much in common and have many differences with the object in the other group. Apart from that most of the data collected in many of the problems seem to have some inherent properties that experienced natural groupings (Hammuda and Karay: 2003). Grouping customers with clustering aided by K-means algorithm. K-means algorithm can perform grouping documents in large numbers. (Ernie & Maria: 2014). K-means is an algorithm to classify or categorize the object object (in this case the data) based on certain parameters into a group, so it can run faster than the hierarchical clustering (if small k) by the number of variables and generates a cluster more tightly , (Budi Susanto: 2010).

1.1 Data Mining

Data mining , often also called knowledge discovery in databases (KDD) , is an activity that includes the collection , use historical data to determine the regularities , patterns or relationships in large data sets . Output of data mining can be used to improve decision making in the future. (B.Santosa : 2007) .

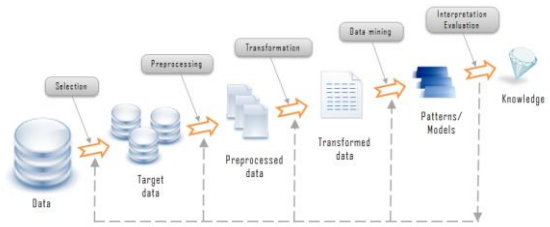
Data mining is a series of processes for adding additional value of a set of data in the form of knowledge that had been unknown manually . (F.A. Hermawati : 2013) .

Data mining is a series of processes for adding additional value of a set of data in the form of knowledge that had been unknown manually , also be interpreted as an analysis automatically from the data for the large amounts or complex in order to find patterns or trends that are important are usually not aware of its existence .

1.2 Stages Data Mining

The term data mining and knowledge discovery in databases (KDD) are often used interchangeably to describe the process of extracting information hidden in a large data base . Actually, the two words have a different concept but relate to each other. And one step in the overall process of KDD is data mining . (Fayyad : 1996) .

KDD process can be broadly described as follows:



Picture 1. Stages KDD

There are several stages in data mining, among others :

1. Selection
   a. Creating a target data set , selecting a subset of data , or focusing on a subset of variables or data samples , where the invention (discovery) will do.
   b. Selection (selection) of data from the operational data set needs to be done before the stage of extracting information in KDD starts. Data from the selection that will be used for data mining process , is stored in a file separate from operational databases .
2. Preprocessing
   a. Preliminary processing and data cleansing are basic operations such as noise removal done .
   b. Before the data mining process can be implemented , it needs cleaning process on the data to be the focus of KDD .
   c. Cleaning process may include removing duplicates the data , check the data inconsistent , and correct errors in the data , such as print errors (typography) .
   d. Do enrichment process , ie the process of " enriching " existing data with the data or other information that is relevant and necessary for KDD , such as data or external information .
3. Transformation
   a. Search features that are useful to present the data depends on the goal to be achieved .
   b. Is the process of transformation on the data that has been selected, so that data corresponding to the data mining process . This process is the creative process and is highly dependent on the type or pattern of information to be searched in the database
4. Data Mining
   a. Selection of data mining tasks ; election goal of KDD process instance classification , regression , clustering , etc .
   b. Selection of data mining algorithms to search (searching)
   c. Process Data mining is the process of searching for a pattern or interesting information in the selected data by using techniques or methods. Techniques, methods , or data mining algorithms in highly variable . Selection of the appropriate method or algorithm relies heavily on the goals and overall KDD process.
5. Evaluation
   a. Translation patterns generated from data mining .
   b. Pattern information generated from the data mining process needs to be presented in a form easily understood by stakeholders
   c. This stage is part of the KDD process that includes checking whether a pattern or information found contradicted by the facts or hypotheses that existed before.

1.3 Method Of Clustering

Clustering grouping of data in the absence of specific data classes , so that the distribution of all data into groups that have similar attributes. Clustering is a grouping of records , observations , or noticed and form a class of objects that have similar (B.Santosa : 2007) .

Cluster analysis which found the object as a collection of objects in the same group (or linked)

with each other and different (or not dealing) with the objects in other groups.

The goal of cluster analysis is to minimize the distance in the cluster and to maximize the distance between clusters . (F.A. Hermawati : 2003).

The main objective of the cluster is a method of grouping a number of data / object into a cluster (group) so that in each cluster will contain the data as closely as possible . In the clustering we are trying to put similar objects (short distance) in one cluster and create a distance between clusters as far as possible .

1.4 Algoritma K-Means

K - Means clustering algorithm is a distance - based cluster techniques that seek to partition the data into several clusters. This method of partitioning the data into clusters according to the characteristics of any data, any data that has the same characteristics are grouped into the same cluster as well as the data possessed different characteristics grouped in another cluster .

In this algorithm , which became the center of the cluster is called the centroid , the centroid is a random value of the entire data set is selected at an early stage , then the K -Means selecting each component of the entire data and separating data into one centroid that has been described previously by the shortest distance between the components of the data center and each centroid provided that no more data to move the group .

Grouping data using K -means :

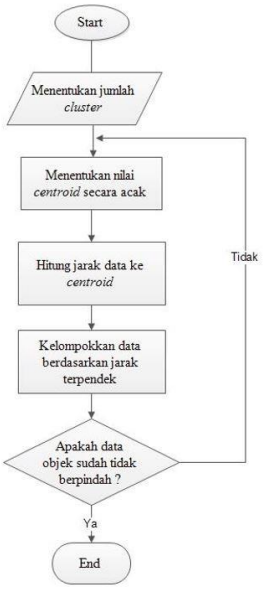
1. Determine the number of clusters
2. Determining the value of a centroid in each cluster . To determine the value of the centroid of the initial iteration is done randomly .
3. Calculate rarely any data at each centroid using the formula Euclidean distance to find the closest to the centroid of any data , the following formula:

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^{n} (x_j - c_j)^2}$$

d = distance
j = the number of data
c = centroid
x = Data

4. Grouping objects to determine the cluster members are taking into account the minimum distance of the object.
5. Go back to step 2 , if there is still data to move the group , until no more data to move the group .
6. Selection Show Results

Flowchart of the algorithm K -Means:



Picture 2. Flowchart K-Means

2. RESEARCH

2.1. Analisys System

The analysis system can be defined as the decomposition of a system whose whole into its component parts with a view to identify and evaluate the problems - problems that occur and needs - the expected demand . In this section will be explained in more detail about the problems and needs to be achieved .

2.1.1. Problem Analisys

Based on the results of field studies that have been done in the pre - research study that included interviews , observation and data collection has been done , so he found some problems that exist in the online division Biobses Book Store , among them:

a. Provision of additional discounts that are difficult to implement.

In the e-commerce system biobses granting extra discounts difficult to implement because of the number of transactions and more each day, so data transaksipun lot. Resulting in less effective due to the results achieved are still in manual calculation. Chronology of giving a discount on bookstore biobses at this present moment in a system that is running is the provision of initial discount already established company already available in their e-commerce, but for an additional discount given in the following months these tasks are still done manually by division their online. Namely in the form of user input is done. With the implementation of this additional discount will be difficult for the online

division of their own because of the number of transactions that many per day. Rules that have been created for the online division should remain implemented as it has been instructed by the company. Calculation of monthly transactions with individual customers is done by one employee is certainly going to take a long time, and less productive

Giving extra discounts are ongoing at the present time is the division of three groups of customer transaction data. In the probable month - next month or next year from its online division biobses parties will amend the policy by giving an extra discount shopping more or minimal. For the settlement of this problem in the system that will be made will be given a choice of cluster formation and selection of discount you want to use by the online division in Biobses.

With details of an additional discount given by the company, the calculation of customer transactions and the number of customers who have a lot of difficulties for the online division, because it must calculate, mengolompokkan and provide discounts to their respective customers.

Calculations should be performed on the last date of each month, which will spend a lot of time and energy in that customer information will get additional discounts to be delivered the next day and delivered via sms media, where in addition to spending time, pulse sms will be a lot of unused

.

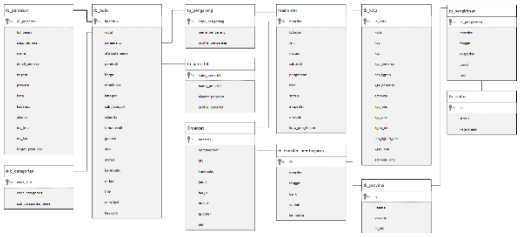
b. Active customer information to companies

Website administrators find it difficult to see which user shopping transactions reached the minimum target and user anywhere will get an extra discount in the following month . Each month the company has a target that is planned to be achieved by their online division melaluik e - commerce itself . With the calculation of transaction data that is still the target is less accurate manual pencapainya in because small mistakes will be found errors in the calculations

.

2.1.2 Analisis Sumber Data

To make the process of data mining first step you should do is look at the source of the data contained within the database system itself . The data will be used is derived from one of the data which is attached to the e - commerce dialam bi - obses . The following relation scheme contained in the e - commerce system biobses :



Picture 3. Relation Scheme Data Source

2.1.3 Impelemtation Of Data Mining

The implementation process of data mining adapted to the KDD process phases :

a. Selection

Early in the KDD process step is the selection of data. The data will be captured from multiple database tables contained in biobses . The data will be used is the transaction data (Table Main Order) , the data subscriber (Table Buyer) , and data transaction details (Order Details table).

Table 1. Tabel Main Order

Nama Field	Tipe Data
noorder	int
tglorder	varchar
jam	time
nocust	varchar
subtotal	decimal
pengiriman	decimal
total	decimal
status	varchar
ekspedisi	varchar
metode	varchar
kota_pengiriman	varchar

Table 2. Tabel Pemesan.

Nama Field	Tipe Data
ID_pemesan	int
Tgl_pesan	Varchar
Pass_pemesan	Varchar
Nama	Varchar
Email	Varchar
Negara	Varchar
Provinsi	Varchar
Kota	Varchar
Kodepos	Varchar
Alamat	Text
No._telp	Varchar
No._fax	Varchar
Forget_pass_key	Varchar

Tabel 3. Tabel Detail Order

Nama Field	Tipe Data
noorder	int
tglorder	int
qty	int

kodebuku	varchar
judul	varchar
harga	decimal
nocust	varchar
tglorder	varchar
jam	time

There are several attributes of each table will be selected . Selection of this attribute refers to the ultimate goal of this research . The following attributes that will be taken from each table .

1. Attributes are taken from Table Main Order
   a. No Order
      These attributes are selected and positioned in the first place because of this attribute is a unique number , No. This order is a purchase ID , selected bedasarkan need to indicate at the time when the transaction occurs .
   b. Total (Rp)
      This attribute is used to view the total customer shopping and to check in to which group the customer.
   c. Status :
      This attribute is used to view the status of customers shopping , where here there are two types , pending and delivered .

2. Atribut yang diambil dari **Tabel Pemesan**
   a. Nama
      These attributes have been selected because the name of the buyer is an attribute that will be used later to see customers with registered names that get additional discounts .
   b. Email
      This attribute is used later for the submission of additional information to the user discounts

3. The following attributes that will be taken from the **Tabel Detail Order.**
   Qty is an attribute number of customers shopping quantity . Bedasarkan been to see the number of customers shopping quantity
   .
   Selection of some of these attributes to produce a single new table . Namely , Table Transaction

.Tabe 4. Tabel Transaksi

No	No. Order ID	Total (Rp)	Status	Nama	Email	
1	10422	104.000	Delivered	M. Ebert Banurea	b.rea@yahoo.com	
2	10420	158.000	Delivered	Iskandar	batamwork@gmail.com	
3	10419	95.000	Pending	Agus Mustopa	agus.mustopa@gmail.com	
4	10418	57.500	Delivered	Deny Wanda	denywanda99@gmail.com	
5	10417	87.000	Delivered	Eldi Winansyah	Eldywin@yahoo.co.id	
6	10415	100.000	Pending	Muhammad Fathul Aziz	adibrojab@gmail.com	
7	10414	50.000	Delivered	Siti Aida Siswarana	staida24@gmail.com	
8	10413	50.000	Delivered	Acun Rukmana	acunrukmana@gmail.com	
9	10412	61.000	Pending	khoirul	khoirullukman1@gmail.com	1
10	10408	70.000	Delivered	Moch Fadholi	moch.fadholi@gmail.com	1
11	10407	73.000	Delivered	tia saptiani	tia.saptiani@gmail.com	1
12	10406	34.000	Delivered	Supri	supri6565@yahoo.co.id	1
13	10403	496.000	Delivered	jordi iswantoro	jordi.iswantoro@yahoo.co.id	7
14	10398	98.000	Delivered	Roy Hiefli	pasualismed@gmail.com	1
15	10393	208.000	Delivered	suryo handoko	sryhandoko@yahoo.com	2
16	10380	39.000	Delivered	Faridudin Arrido	faridudin.arrido@gmail.com	1
17	10372	70.000	Delivered	Sanusi	sanusi.spd.mkom@gmail.com	1
18	10370	121.150	Delivered	Denny Halim	d3nnyh4l1m@gmail.com	1
19	10373	129.000	Pending	Johnie Pasaribu	jrspasaribu@gmail.com	1
20	10389	122.000	Delivered	Yana Setiono	yans.start@gmail.com	1
21	10390	98.000	Delivered	Fahrur Rozi	mailrozi@gmail.com	1
22	10385	62.000	Pending	Nurmansyah	oklahklobegitu@gmail.com	1
23	10371	127.000	Delivered	Maharani Sekar Melati	melati.96@gmail.com	1
24	10375	104.000	Delivered	Muhammad Ali Mursid Alfathoni	mhd.ali8mursid@gmail.com	1
25	10400	100.000	Delivered	sulistyawan	mas.sulistyawan@gmail.com	1

b. Preprocessing

After making the selection process the data that is to get some table to be used next process is the cleaning and selection.

1. Cleaning

Cleaning is the process of removing or cleaning some data that is not in accordance with the data requirements for the next process , namely the data selection process .

The process that will be done is the deletion of data whose status is still pending . Pending status means that a customer is a transaction is not completed or failed .

2. Selection

At this stage of the selection process will be conducted for data processing miningnya , the selected attributes to KDD process are :

a. Total (Rp)
   This attribute is used to view the total customer shopping and to check in to which group the customer.
b. Qty
   Qty is an attribute number of customers shopping quantity . Selected by to see the number of customers shopping quantity .

c. Data Mining

After going through several steps above , at this stage will mmebahas miningnya data processing itself . In this study the existing problem using clustering or grouping method possible

The aim is to establish a data group (cluster) of a set of data that is not known to their groups or the classes and to determine the data included in the cluster which.

This analysis phase authors use k-means algorithm . This algorithm uses to create a cluster centroid . Centroid is also used to calculate the distance of an object's data to the centroid . A data objects included in a cluster if it has the shortest distance to the cluster centroid .

Step -step operation :
Is known :
Total Cluster = 3
Total data = 20
Attribute Number = 2

Attributes selected from the results table dry , the total attribute (rp) and attributes qty .

Thus , the calculation of the K -Means clustering will use attributes total (rp) and attribute qty , attributes the total transaction been due to see the entrance to which group the data, while the attribute qty been selected for this attribute will be a guideline of how much the number of books sold in a single transaction

.

Tabel 5. Tabel Hasil Selection

No	Total (Rp)	Qty
1	104.000	2
2	158.000	2
3	57.500	1
4	87.000	1
5	50.000	1
6	50.000	1
7	70.000	1
8	73.000	1
9	34.000	1
10	496.000	7
11	98.000	1
12	208.000	2
13	39.000	1
14	70.000	1
15	121.150	1
16	122.000	1
17	98.000	1
18	127.000	1
19	104.000	1
20	100.000	1

The next step is to look for patterns to be found for grouping data mining transaction data . Using the k - means algorithm . Next step - its algorithm steps

1. Determine the number of clusters
   The number of clusters to be selected is as much as three clusters
2. Determining the value of a centroid in each cluster To determine the value of the centroid of the initial iteration is done randomly
   - Taken from the data to - 9
     Total Books : 1
     Total transaction : 34,000
   - Taken from the data to - 12
     Total Books : 2
     Total Transactions : 208,000
   - Taken from the data to - 10
     Total Books : 7
     Total Transactions : 496.000

Table. Cluster :

Centroid	X1	X2
C_1	1	34000
C_2	2	208000
C_3	7	496000

Ket :
X1 = Total Buku
X2 = Total Transaksi

3. Calculate the distance of each data to the nearest centroid . The closest cluster centroid will be followed by the data.

Distance data by cluster 1 is :

$$d(x_1, c_1) = \sqrt{(x1_1 - c_{1a})^2 + (x2_1 - c_{1b})^2}$$
$$= \sqrt{(2-1)^2 + (104.000 - 34.000)^2}$$
$$= \sqrt{(1)^2 + (70.000)^2}$$
$$= 70.001$$

$$d(x_2, c_1) = \sqrt{(x1_2 - c_{1a})^2 + (x2_2 - c_{1b})^2}$$
$$= \sqrt{(2-1)^2 + (158.000 - 34.000)^2}$$
$$= \sqrt{(1)^2 + (124.000)^2}$$
$$= 124.001$$

$$d(x_3, c_1) = \sqrt{(x1_3 - c_{1a})^2 + (x2_3 - c_{1b})^2}$$
$$= \sqrt{(1-1)^2 + (57.500 - 34.000)^2}$$
$$= \sqrt{(0)^2 + (23.500)^2}$$
$$= 23.500$$

Perform calculations on each cluster and the centroid ,so the result was found :

Table 7. Calculation Results

No	Total Buku	Total (Rp)	Dc1	Dc2	Dc3
1	2	104.000	70.000	104.000	392.000
2	2	158.000	124.000	50.000	338.000
3	1	57.500	23.500	150.500	438.500
4	1	87.000	53.000	121.000	409.000
5	1	50.000	16.000	158.000	446.000
6	1	50.000	16.000	158.000	446.000
7	1	70.000	36.000	138.000	426.000
8	1	73.000	39.000	135.000	423.000
9	1	34.000	0	174.000	462.000
10	7	496.000	462.000	288.000	0
11	1	98.000	64.000	144.000	398.000
12	2	208.000	174.000	0	288.000
13	1	39.000	5000	169.000	457.000
14	1	70.000	36.000	138.000	426.000
15	1	121.150	87.150	86.850	374.500
16	1	122.000	88.000	86.000	374.000
17	1	98.000	64.000	110.000	398.000
18	1	127.000	93.000	81.000	369.000
19	1	104.000	70.000	104.000	392.000
20	1	100.000	66.000	108.000	396.000

4. Segment data according to the cluster , the data that has the shortest distance

Table. 8. Calculation Results spacing and grouping data

No.	Total Buku	Total (Rp)	DC1	Dc2	Dc3	C1	C2	C3
1	2	104.000	70.000	104.000	392.000	*		
2	2	158.000	124.000	50.000	338.000		*	
3	1	57.500	23.500	150.500	438.500	*		
4	1	87.000	53.000	121.000	409.000	*		
5	1	50.000	16.000	158.000	446.000	*		
6	1	50.000	16.000	158.000	446.000	*		
7	1	70.000	36.000	138.000	426.000	*		
8	1	73.000	39.000	135.000	423.000	*		
9	1	34.000	0	174.000	462.000	*		
10	7	496.000	462.000	288.000	0			
11	1	98.000	64.000	144.000	398.000	*		
12	2	208.000	174.000	0	288.000		*	
13	1	39.000	5000	169.000	457.000	*		
14	1	70.000	36.000	138.000	426.000	*		
15	1	121.150	87.150	86.850	374.500		*	
16	1	122.000	88.000	86.000	374.000		*	
17	1	98.000	64.000	110.000	398.000	*		
18	1	127.000	93.000	81.000	369.000		*	
19	1	104.000	70.000	104.000	392.000	*		
20	1	100.000	66.000	108.000	396.000	*		

5. Calculate the new Cluster Center

Furthermore, a new centroid is calculated for each cluster based on data that is joined on each cluster.

- For the First Cluster

For the first cluster there are 14 data belonging to it

.

Table 8 Cluster 1

Data	Fitur a	Fitur b
1	2	104.000
3	1	57.500
4	1	87.000
5	1	50.000
6	1	50.000
7	1	70.000
8	1	73.000
9	1	34.000
11	1	98.000
13	1	39.000
14	1	70.000
17	1	98.000
19	1	104.000
20	1	100.000
Nk	Jumlah a	Jumlah b
14	15	1.034.500
Rata-Rata	1.07	73.892

- For the Second Cluster

For the second cluster , there are 7 data incorporated therein

Table 9. Cluster 2

Data	Fitur a	Fitur b
2	2	158.000
12	2	208.000
15	1	121.150
16	1	122.000
18	1	127.000
Nk	Jumlah a	Jumlah b
5	7	736.150
Rata-Rata	1,4	147.230

- For Third Cluster

For the third cluster , there is one data incorporated therein .

Centroid	a	b
1	1,07	73.892
2	1,4	147.230
3	7	496.000

Repeat the process at this stage of the previous calculation using the latest cluster , on this second iteration centroid position there is no change , then the calculation process is terminated , that is, the clustering process has been completed to the end.

d. Knowledge

The patterns formed are three groups of customer transaction data , with details as follows :

a. Discounts 2 , 5 % (Transaction Rp . 0 - Rp . 100,000) There are 14 data customers who have transactions starting from Rp . 34.000 - Rp . 100,000 which will get an additional discount of 2.5 % for next month's transactions .

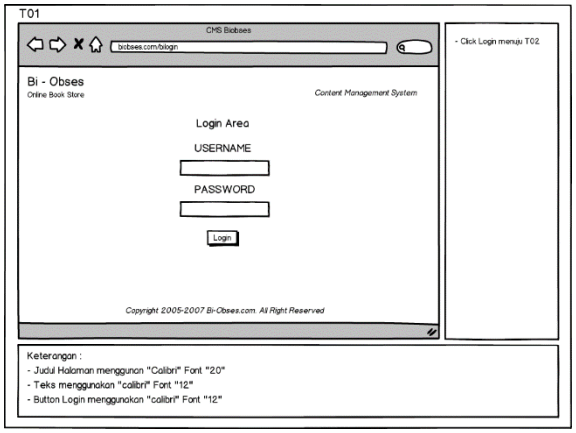
b. 5% discount (Transaction Rp . 110,000 - Rp . 400,000) There are 5 customer data has begun transactions of Rp . 121 150 - Rp . 208,000 which will get an additional discount of 2.5 % for next month's transactions .

c. 10% % (Deal Rp . 410,000 - Rp . 1,000,000) There are 1 data of customers who have initiated a transaction of Rp . 469,000 which will get an additional discount of 10% for next month's transactions .

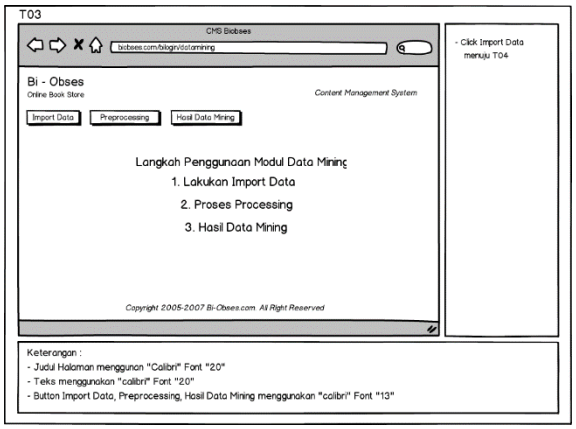
With the advent of customer transaction data sharing in biobses then the data mining process has been completed

3. DESIGN OF INTERFACE

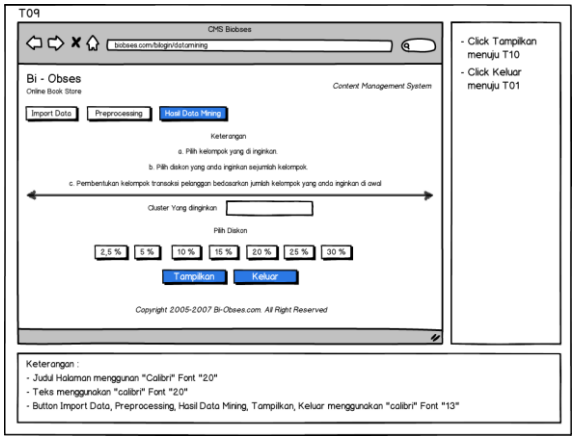
Here is an example of the design of the interface of the device data mining to be built :



Picture 4. Login Page

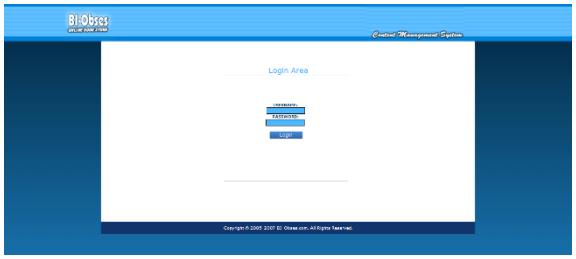


Picture 5. Moduls Of Data Mining



Picture 6. Result Of Data Mining

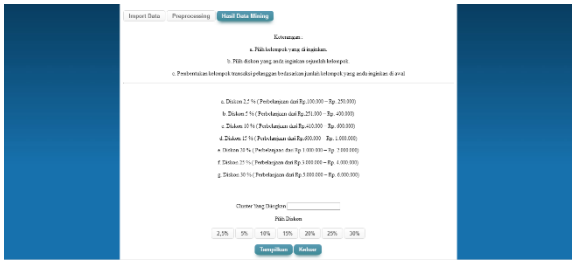
4. IMPLEMENTATION OF INTERFACE



Picture 7. Login Implementation



Picture 8. Implementaton Of Data Mining Page



Picture 9. Implementation of Result Data Mining

5. CONCLUSION

After doing the analysis , design, implementation and its testing has been done , it can be concluded on the application of data mining is that the application is built can help in the Online Division of grouping customers bedasarkan their transactions. The addition of this feature is expected not to be found difficulty in classifying their customer data . So the development of data mining applications is in accordance with the procedures and the expected goals .

6. BIBLIOGRAPHY

[1] A. Kadir, Konsep dan Tuntunan Basis Data, Yogyakarta: Andi, 2001.
[2] Budi Santosa, Data Mining :Teknik Pemanfaatan Data untuk Keperluan Bisnis, 1st ed. Yogyakarta, Indonesia: Graha Ilmu, 2007
[3] F. A. Hermawati, Data Mining, Yogyakarta: ANDI, 2013.
[4] Janner Simarmata and Iman Prayuda, Basis Data. Yogyakarta: Andi Offset, 2006.
[5] Kusrini and E.T. Luthfi, Algoritma Data Mining, Yogyakarta: Andi Publisher, 2009
[6] M. Flowler, UML Distiled Edisi Ketiga Panduan Singkat Tentang Bahasa Permodelan Objek Standar, Yogyakarta: Andi, 2004.
[7] M. Sadeli, Visual Basic 2013, Palembang: Maxicom, 2013
[8] Prasetyo, E. Data Mining : Konsep Dan Aplikasi Menggunakan Matlab. Yogyakarta: Andi Publisher, 2013
[9] Prasetyo, E. Data Mining : Mengolah Data Menjadi Informasi Menggunakan Matlab. Yogyakarta: Andi Publisher, 2014

[10] Raharjo, Budi. Mudah Belajar PHP. Bandung, Indonesia : Informatika:2015
[11] S. Susanto and D. Suryadi, Pengantar Data Mining, Yogyakarta: ANDI, 2010.
[12] Yudi Agusta, "K-Means - Penerapan, permasalahan dan Metode Terkait," Jurnal Sistem dan Informatika, vol. III, pp. 47-60, Februari 2007.