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

IMPLEMENTASI DATA MINING METODE CLUSTERING PADA TOKO BUKU BI-OBSES BANDUNG DENGAN PENERAPAN ALGORITMA K-MEANS

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 *bookstore* adalah toko buku komputer yang telah menerapkan sistem *e-commerce* dalam kegiatan transaksi jual belinya. Pada sistem ini juga diterapkan sistem pemberian diskon untuk setiap transaksi pelanggan mereka. Sistem *e-commerce* yang ada saat ini mengalami kesulitan pada proses pemberian diskon khususnya diskon tambahan untuk pelanggan, hal ini disebabkan oleh tidak adanya pengelompokan pelanggan berdasarkan jumlah transaksi yang mereka lakukan setiap bulannya.

Metode *clustering* dalam *data mining* merupakan solusi yang tepat untuk mengatasi permasalahan pengelompokkan data pelanggan. Dengan menerapkan algoritma *k-means* pembentukan kelompok pelanggan dimulai dengan menentukan jumlah *cluster* yang akan di bentuk, menentukan nilai *centroid* masing-masing *cluster*, menghitung jarak antar data, dan mengelompokkan dengan memperhitungkan jarak minimum objek.

Melalui aplikasi *data mining* dengan metode *clustering* proses pengelompokan data pelanggan pada sistem *e-commerce* Biobses *bookstore* menjadi lebih mudah. Selain itu, permasalahan pada pemberian diskon tambahan dapat diatasi dan menjadi lebih mudah dengan sistem diskon yang bisa di *customize* setiap bulannya oleh perusahaan.

Kata Kunci : Biobses, C*lustering*, *Data Mining*, Pengelompokan,

1. PENDAHULUAN

Biobses merupakan sebuah perusahaan yang bergerak dibidang penjualan buku-buku komputer dari skala kecil hingga menengah dengan pasar yang tersebar di seluruh Indonesia. Bi-obses sudah menggunakan sistem *E-commerce* untuk pemasaran dan penjualan produknya. Sistem *E-Commerce* ini dikelola oleh Divisi Online dengan tugas utama yaitu mengatur transaksi online, promosi dan pemberian diskon.

Ditoko buku biobses memiliki dua alur proses bisnis yaitu offline dan online. Di kedua proses bisnis ini biobses menetapkan pemberian diskon yang sudah di batasi oleh perusahaan. Perbedaanya adalah di alur bisnis transaksi offline, ketika pelanggan melakukan transaksi dan kebetulan jumlah belanja minimal tercapai akan diberikan diskon tambahan oleh perusahaan. Akan tetapi di online sistem yang di terapkan perusahaan berbeda, dimana user akan mendapatkan diskon tambahan melalui transaksi mereka selama sebulan, jika kuota perbelanjaan tercapai, dan diberikan diskon tambahan pada bulan berikutnya. Akan tetapi pemberian diskon ini sulit diterapkan karena transaksi online di biobses yang banyak perharinya. Sehingga informasi diskon tambahan yang diberikan kepada pelanggan mereka sendiri tidak bisa tersampaikan pada setgiap akhir bulanya. Dan target awal pemasaran yaitu menemukan pelanggan sebanyak mungkin yang ditugaskan pada divisi online pun tidak tercapai di karenakan sulitnya untuk mengelompokkan pelanggan bedasarkan transaksi mereka.

Salah satu upaya untuk menyelesaikan permasalahan pemberian diskon ini adalah dengan melakukan pengelompokkan pelanggan dengan menggunakan prinsip data mining. Dengan melakukan data mining, perusahaan dapat membentuk pengelompokkan pelanggan bedasarkan transaksi yang telah mereka lakukan dan menganalisa pasar untuk bisa memahami pasar yang sudah ada ataupun menemukan peluang-peluang yang baru untuk meningkatkan keuntungan termasuk memprediksi kriteria pemasaran. Teknologi ini sekarang sudah ada dan di aplikasikan oleh perusahaan-perusahaan untuk memecahkan berbagai permasalahan bisnis (Achmad Zulfikar : 2009). *Data mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies : 2004). Salah satu metode *data mining* yaitu *clustering* adalah proses pengelompokkan kumpulan data menjadi beberapa kelompok sehingga objek di dalam satu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaan dengan objek dikelompok lain. Selain dari itu sebagian besar data yang dikumpulkan dalam banyak masalah

terlihat memiliki beberapa sifat yang melekat yang mengalami pengelompokan-pengelompokan natural (Hammuda dan Karay : 2003). Pengelompokkan pelanggan dengan clustering di bantu dengan algoritma *K-means*. Algoritma *K-means* dapat melakukan pengelompokkan dokumen dalam jumlah yang banyak. (Ernie & Maria : 2014). *K-means* adalah sebuah algoritma untuk mengklasifikasikan atau mengelompokkan objek objek (dalam hal ini data) berdasarkan parameter tertentu ke dalam sejumlah group, sehingga dapat berjalan lebih cepat daripada hierarchical *clustering* (jika k kecil) dengan jumlah variable yang besar dan menghasilkan *cluster* yang lebih rapat. (Budi Susanto : 2010).

1.1 Data Mining

Data mining, sering juga disebut *knowledge discovery in database (KDD)*, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menentukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. *(B.Santosa : 2007).*

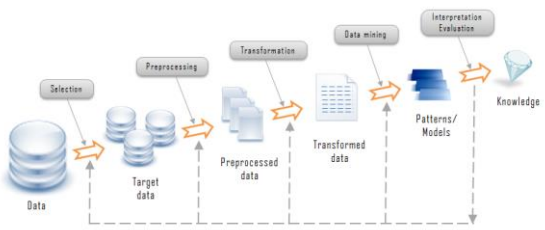
Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. *(F.A. Hermawati : 2013).*

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual, juga diartikan sebagai analisa otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya.

1.2 Tahapan Data Mining

Istilah data mining dan *knowledge discovery in databases (KDD)* sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda akan tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. (*Fayyad : 1996).*

Proses KDD secara garis besar dapat dijelaskan sebagai berikut :



Gambar 1. Tahapan KDD

Terdapat beberapa Tahapan dalam data mining, antara lain :

1. Selection
   a. Menciptakan himpunan data target , pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (discovery) akan dilakukan.
   b. Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.
2. Preprocessing
   a. Pemprosesan pendahuluan dan pembersihan data merupakan operasi dasar seperti penghapusan noise dilakukan.
   b. Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD.
   c. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).
   d. Dilakukan proses enrichment, yaitu proses "memperkaya" data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.
3. Transformation
   a. Pencarian fitur-fitur yang berguna untuk mempresentasikan data bergantung kepada goal yang ingin dicapai.
   b. Merupakan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data
4. Data Mining
   a. Pemilihan tugas data mining; pemilihan goal dari proses KDD misalnya klasifikasi, regresi, clustering, dll.
   b. Pemilihan algoritma data mining untuk pencarian (searching)
   c. Proses Data mining yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung

pada tujuan dan proses KDD secara keseluruhan.

5. Evaluation
   a. Penerjemahan pola-pola yang dihasilkan dari data mining.
   b. Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.
   c. Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

1.3 Metode Clustering

Clustering melakukan pengelompokan data tanpa adanya kelas data tertentu, sehingga dilakukan pembagian seluruh data menjadi kelompokkelompok yang memiliki kemiripan atribut. Pengklusteran merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan *(B.Santosa : 2007).*

Analisa cluster yaitu menemukan kumpulan objek hingga objek-objek dalam satu kelompok sama (atau punya hubungan) dengan yang lain dan berbeda (atau tidak berhubungan) dengan objek-objek dalam kelompok lain. Tujuan dari analisa cluster adalah meminimalkan jarak di dalam cluster dan memaksimalkan jarak antar cluster. *(F.A. Hermawati : 2003).*

Tujuan utama dari metode klaster adalah pengelompokan sejumlah data/objek ke dalam klaster(group) sehingga dalam setiap klaster akan berisi data yang semirip mungkin. Dalam clustering kita berusaha untuk menempatkan obyek yang mirip (jaraknya dekat) dalam satu klaster dan membuat jarak antar klaster sejauh mungkin.

1.4 Algoritma K-Means

Algoritma K-Means clustering merupakan teknik cluster berbasis jarak yang berusaha mempartisi data kedalam beberapa cluster. Metode ini mempartisi data kedalam cluster menurut karakteristik yang dimiliki setiap data, setiap data yang memiliki karakteristik sama dikelompokkan kedalam satu cluster yang sama begitu juga dengan data yang mempunya karakteristik berbeda dikelompokkan kedalam cluster lain .

Pada algoritma ini, yang menjadi pusat cluster dinamakan centroid, centroid merupakan nilai acak dari seluruh kumpulan data yang dipilih pada tahap awal, kemudian K-Means menyeleksi masing-masing komponen dari seluruh data dan memisahkan data tersebut kedalam salah satu centroid yang sudah diuraikan sebelumnya berdasarkan jarak terdekat antara komponen data dan pusat masing masing centroid dengan syarat tidak ada lagi data yang berpindah kelompok.

Pengelompokan data menggunakan K-means :

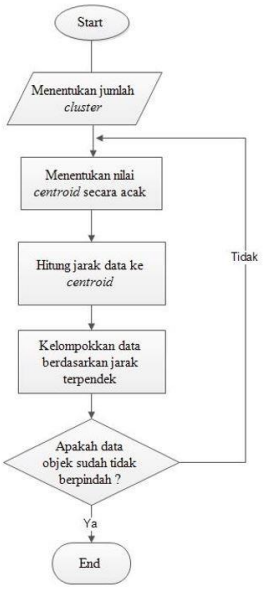
1. Menentukan jumlah cluster
2. Menentukan nilai centroid yang ada di masing-masing cluster. Untuk menentukan nilai centroid pada awal iterasi dilakukan secara acak.
3. Hitung jarang setiap data yang ada pada masing-masing centroid menggunakan rumus euclidian hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid, berikut rumusnya :

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

d = jarak
j = banyaknya data
c = centroid
x = data

4. Pengelompokan objek untuk menentukan anggota cluster adalah dengan memperhitungkan jarak minimum objek.
5. Kembali ke langkah 2, apabila masih ada data yang berpindah kelompok, sampai tidak ada lagi data yang berpindah kelompok.
6. Tampilkan Hasil seleksi

Flowchart dari algoritma K-Means :



Gambar 2. Flowchart K-Means

2. ISI PENELITIAN

2.1. Analisis Sistem

Analisis sistem dapat didefinisikan sebagai penguraian dari suatu sistem yang utuh ke dalam bagian-bagian komponennya dengan maksud untuk mengidentifikasi dan mengevaluasi permasalahan-permasalah yang terjadi dan kebutuhan – kebutuhan yang diharapkan. Pada bagian ini akan dijelaskan lebih detail mengenai permasalahan serta kebutuhan yang ingin dicapai.

1.1.1 Analisis Masalah

Berdasarkan hasil studi lapangan yang telah dilakukan dalam penelitian pre-riset yang meliputi wawancara, observasi dan pengumpulan data yang telah dilakukan, maka didapatkan beberapa masalah yang ada pada divisi online Biobses Book Store, di antaranya :

a. Pemberian diskon tambahan yang sulit di terapkan.

Dalam system e-commerce biobses pemberian diskon tambahan sulit diterapkan dikarenakan jumlah transaksi yang banyak setiap harinya, sehingga data transaksipun banyak. Mengakibatkan kurang efektifnya hasil yang dicapai dikarenakan masih dalam perhitungan manual.

Alur pemberian diskon pada toko buku biobses pada saat sekarang ini dalam system yang berjalan adalah pemberian diskon awal yang sudah ditetapkan perusahaan sudah tersedia dalam e-commerce mereka, akan tetapi untuk diskon tambahan yang di berikan pada bulan berikutnya tugas ini masih dilakukan secara manual oleh divisi online mereka. Yaitu berupa inputan manual yang dilakukan.

Dengan penerapan diskon tambahan ini akan menjadi kesulitan sendiri bagi divisi online mereka dikarenakan jumlah transaksi yang banyak perharinya. Peraturan yang telah di buat untuk divisi online harus tetap terlaksana sebagaimana yang telah di perintahkan oleh perusahaan. Dengan perhitungan transaksi perbulan secara individu pelanggan dikerjakan oleh satu orang karyawan ini tentunya akan memakan waktu yang lama, dan kurang produktif.

Pemberian diskon tambahan yang sedang berjalan pada saat sekarang ini adalah pembagian 3 kelompok data transaksi pelanggan. Pada Besar kemungkinan pada bulan - bulan berikut atau tahun berikutnya pihak dari divisi online biobses akan merubah kebijakan dengan pemberian diskon tambahan yang lebih atau minimal perbelanjaan. Untuk penyelesaian masalah ini di sistem yang akan dibuat akan diberikan pilihan pembentukan cluster dan pilihan diskon yang ingin digunakan oleh pihak divisi online di Biobses.

Dengan rincian diskon tambahan yang diberikan perusahaan, perhitungan transaksi pelanggan dan jumlah pelanggan yang banyak ini merupakan kesulitan bagi divisi online, karena harus menghitung, mengolompokkan dan memberikan diskon ke masing masing pelanggan mereka.

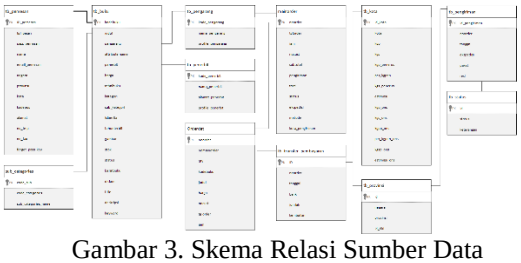
Perhitungan harus dilakukan pada tanggal terakhir setiap bulanya, dimana akan menghabiskan banyak waktu dan tenaga dimana informasi pelanggan yang akan mendapatkan diskon tambahan harus disampaikan keesokan harinya dan disampaikan lewat media sms, dimana selain menghabiskan waktu, pulsa sms pun akan banyak terpakai.

b. Informasi pelanggan aktif ke perusahaan

Administrator website menemukan kesulitan untuk melihat user mana yang transaksinya mencapai target minimum perbelanjaan dan user mana saja yang akan mendapatkan diskon tambahan pada bulan berikutnya. Setiap bulanya perusahaan memiliki target yang direncakan agar di capai oleh divisi online mereka melaluik e-commerce itu sendiri. Dengan perhitungan data transaksi yang masih manual target tersebut kurang akurat pencapainya di karenakan akan ditemukan kesalahan kesalahan kecil dalam perhitungan.

1.1.2 Analisis Sumber Data

Untuk melakukan proses data mining langkah pertama yang harus dilakukan adalah melihat sumber data yang terdapat didalam sistem database itu sendiri. Data yang akan digunakan berasal dari salah satu data yang tedapat dialam e-commerce bi-obses. Berikut skema relasi yang terdapat dalam sistem e-commerce biobses :



Gambar 3. Skema Relasi Sumber Data

1.1.3 Implementasi Data Mining

Proses implementasi data mining disesuaikan dengan tahapan Proses KDD :

a. Selection

Proses awal dalam langkah KDD adalah pemilihan data. Data yang akan dipilih diambil dari beberapa tabel yang terdapat di database

biobses. Data yang akan digunakan adalah *data transaksi* (Tabel Main Order), *data pemesan* (Tabel Pemesan), dan *data detail transaksi* (Tabel Detail Order).

Tabel 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

Tabel 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

Ada beberapa atribut dari masing masing tabel yang akan dipilih. Pemilihan atribut ini mengacu kepada tujuan akhir dari penelitian ini. Berikut atribut yang akan di ambil dari masing-masing tabel.

1. Atribut yang akan di ambil dari **Tabel Main Order** :
   a. No Order
      Atribut ini dipilih dan di posisikan di tempat pertama dikarenakan atribut ini merupakan sebuah nomor unik, No order ini merupakan ID pembelian, dipilih bedasarkan kebutuhan untuk menandakan pada saat kapan proses transaksi terjadi.
   b. Total (Rp)
      Atribut ini digunakan untuk melihat total perbelanjaan pelanggan dan untuk memeriksa masuk ke kelompok manakah pelanggan tersebut.
   c. Status :
      Atribut ini digunakan untuk melihat status perbelanjaan pelanggan, dimana disini terdapat dua jenis yaitu, pending dan delivered.

2. Atribut yang diambil dari **Tabel Pemesan**
   a. Nama
      Atribut ini dipilih dikarenakan nama pemesan merupakan atribut yuang akan digunakan nantinya untuk melihat pelanggan dengan nama terdaftar yang mendapatkan diskon tambahan.
   b. Email
      Atribut ini digunakan nantinya untuk penyampaian informasi diskon tambahan kepada user.

3. Berikut atribut yang akan di ambil dari **Tabel Detail Order.**
   Qty merupakan atribut jumlah quantity perbelanjaan pelanggan. Dipilih bedasarkan untuk melihat jumlah quantity perbelanjaan pelanggan.

Pemilihan beberapa atribut ini menghasilkan satu buah tabel baru. Yaitu, Tabel Transaksi.

Tabel 4. Tabel Transaksi

No	No. Order ID	Total (Rp)	Status	Nama	Email	Qty
1	10422	104.000	Delivered	M. Ebert Banurea	b.rea@yahoo.com	2
2	10420	158.000	Delivered	Iskandar	batamwork@gmail.com	2
3	10419	95.000	Pending	Agus Mustopa	agus.mustopa@gmail.com	1
4	10418	57.500	Delivered	Deny Wanda	denywanda99@gmail.com	1
5	10417	87.000	Delivered	Eldi Winansyah	Eldywin@yahoo.co.id	1
6	10415	100.000	Pending	Muhammad Fathul Aziz	adibrojab@gmail.com	1
7	10414	50.000	Delivered	Siti Aida Siswarana	staida24@gmail.com	1
8	10413	50.000	Delivered	Acun Rukmana	acunrukmana@gmail.com	1
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

Setelah melakukan proses seleksi data yaitu dengan mendapatkan beberapa tabel yang akan digunakan Proses selanjutnya adalah cleaning dan selection.

1. Cleaning

Cleaning adalah proses penghapusan atau pembersihan beberapa data yang tidak sesuai dengan syarat data untuk proses selanjutnya, yaitu proses seleksi data.

Proses yang akan dilakukan adalah penghapusan data yang statusnya masih pending. Status pending maksudnya adalah transaksi yang dilakukan pelanggan tidak selesai atau gagal.

2. Selection

Pada tahapan ini akan dilakukan proses seleksi untuk pengolahan data miningnya, atribut yang dipilih untuk melakukan proses KDD adalah :

a. Total (Rp)
Atribut ini digunakan untuk melihat total perbelanjaan pelanggan dan untuk memeriksa masuk ke kelompok manakah pelanggan tersebut.

b. Qty
Qty merupakan atribut jumlah quantity perbelanjaan pelanggan. Dipilih bedasarkan untuk melihat jumlah quantity perbelanjaan pelanggan.

c. Data Mining

Setelah melalui beberapa tahapan diatas, di tahap ini akan mmebahas proses data miningnya itu sendiri. Pada penelitian ini dengan masalah yang ada menggunakan motode *clustering* atau *pengelompokkan*.

Tujuanya adalah membentuk kelompok data (cluster) dari himpunan data yang tidak diketahui kelompok-kelompoknya atau kelas-kelasnya dan untuk menentukan data-data termasuk kedalam cluster yang mana.

Tahap analisis ini penulis menggunakan algoritma k-means. Algoritma ini menggunakan centroid untuk membuat cluster. Centroid juga digunakan untuk menghitung jarak suatu objek data terhadap centroid. Suatu objek data termasuk dalam suatu cluster jika memiliki jarak terpendek terhadap centroid cluster tersebut.

Langkah -langkah pengoperasian :

Diketahui :

Jumlah Cluster = 3

Jumlah Data = 20

Jumlah Atribut = 2

Atribut dipilih dari tabel hasil cleaning, yaitu atribut total (rp) dan atribut qty.

Jadi, perhitungan K-Means clustering akan menggunakan atribut total (rp) dan atribut qty, atribut total transaksi dipilih dikarenakan untuk melihat masuk ke kelompok mana data tersebut, sedangkan untuk atribut qty dipilih karena atribut ini nantinya akan menjadi pedoman seberapa banyak jumlah buku yang terjual dalam sekali transaksi.

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

Langkah selanjutnya adalah mencari pola yang akan ditemukan data mining untuk pengelompokkan data transaksi. Menggunakan algortima k-means. Berikut langkah – langkah algortimanya :

1. Menentukan jumlah cluster
   Jumlah cluster yang akan di pilih adalah sebanyak 3 cluster.
2. Menentukan nilai centroid yang ada di masing-masing cluster. Untuk menentukan nilai centroid pada awal iterasi dilakukan secara acak.
   - Diambil dari data ke - 9
     Total Buku : 1
     Total Transaksi : 34.000
   - Diambil dari data ke – 12
     Total Buku : 2
     Total Transaksi : 208.000
   - Diambil dari data ke – 10
     Total Buku : 7
     Total Transaksi : 496.000

Tabel 6. Cluster :

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

Ket :
X1 = Total Buku
X2 = Total Transaksi

3. Menghitung jarak setiap data ke centroid terdekat. Centroid terdekat akan menjadi cluster yang diikuti oleh data tersebut. Jarak data dengan *cluster* 1 adalah :

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

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

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

Lakukan perhitungan terhadap masing masing cluster dan centroid, sehingga ditemukanlah hasil :

Tabel 7. Hasil Perhitungan

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. Kelompokkan data sesuai dengan cluster, yaitu data yang memiliki jarak terpendek.

Tabel. 8. Hasil Perhitungan jarak dan pengelompokkan 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. Hitung Pusat Cluster baru
Selanjutnya dihitung centroid yang baru untuk setiap cluster bedasarkan data yang bergabung pada setiap clusternya.
   - Untuk Cluster Pertama
     Untuk cluster pertama terdapat 14 data yang tergabung ke dalamnya.

Tabel. 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

   - Untuk Cluster Kedua
     Untuk cluster kedua, ada 7 data yang tergabung di dalamnya.

Tabel. 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

- Untuk Cluster Ketiga
  Untuk cluster ketiga, ada 1 data yang tergabung di dalamnya.

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

Ulangi proses perhitungan pada tahapan sebelumnya menggunakan cluster terbaru, pada iterasi kedua ini posisi centroid tidak ada perubahan, maka proses perhitungan diberhentikan, artinya, proses clustering sudah selesai sampai akhir.

d. Pengetahuan/Knowledge

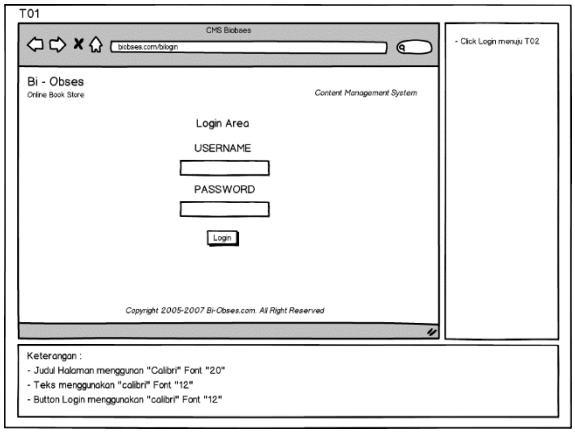
Pola yang terbentuk adalah tiga kelompok data transaksi pelanggan, dengan rincian sebagai berikut :

a. Diskon 2, 5 % (Transaksi Rp. 0 – Rp. 100.000)
   Terdapat 14 data pelanggan yang memiliki transaksi dimulai dari Rp. 34.000 – Rp. 100.000 yang akan mendapatkan diskon tambahan sebesar 2,5 % untuk transaksi bulan berikutnya.
b. Diskon 5 % (Transaksi Rp. 110.000 – Rp. 400.000)
   Terdapat 5 data pelanggan yang memiliki transaksi dimulai dari Rp. 121.150 – Rp. 208.000 yang akan mendapatkan diskon tambahan sebesar 2,5 % untuk transaksi bulan berikutnya.
c. Diskon 10 % % (Transaksi Rp. 410.000 – Rp. 1.000.000)
   Terdapat 1 data pelanggan yang memiliki transaksi dimulai dari Rp. 469.000 yang akan mendapatkan diskon tambahan sebesar 10 % untuk transaksi bulan berikutnya.

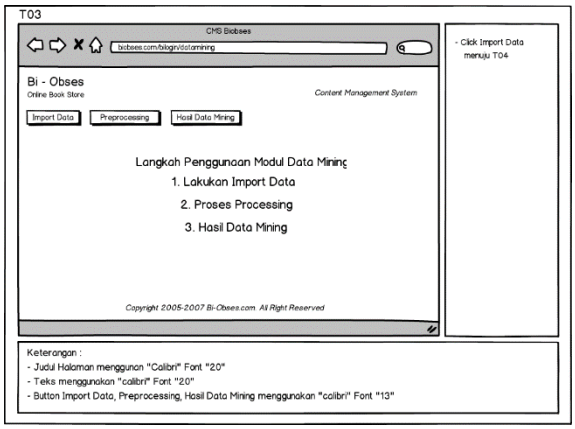
Dengan telah munculnya pembagian data transaksi pelanggan di biobses maka proses data mining telah selesai.

3. PERANCANGAN ANTARMUKA

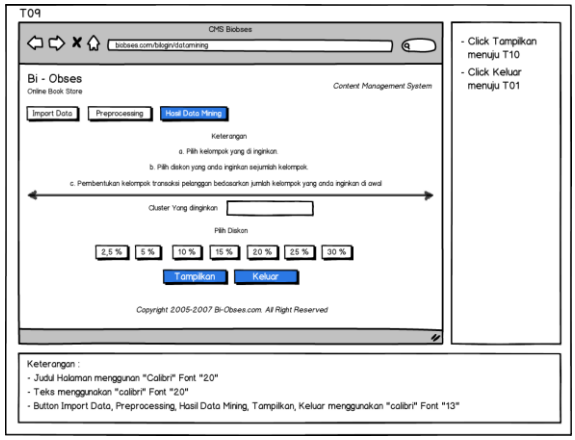
Berikut merupakan contoh perancangan antarmuka dari perangkat *data mining* yang akan dibangun :



Gambar 4. Halaman Login

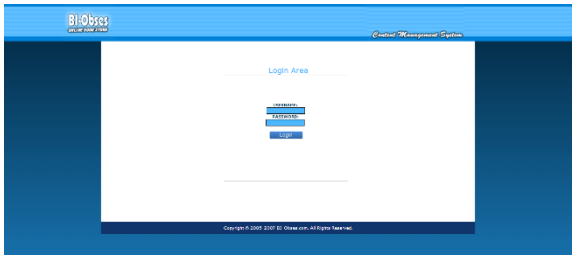


Gambar 5. Data Mining Modul



Gambar 6. Hasil Data Mining

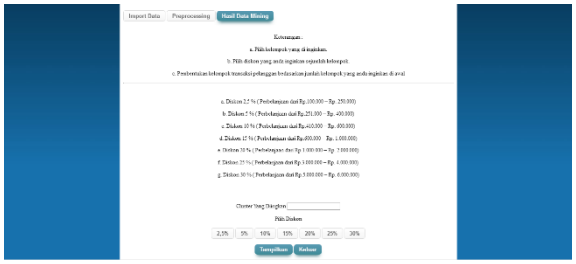
4. IMPLEMENTASI ANTARMUKA



Gambar 7. Impelementasi Login



Gambar 8. Implementasi Halaman Data Mining



Gambar 9. Implementasi Hasil Data Mining

5. KESIMPULAN

Setelah melakukan analisis, perancangan, implementasi beserta pengujian yang telah dilakukan, maka dapat diperoleh kesimpulan terhadap aplikasi *data mining* ini bahwa aplikasi yang di bangun dapat membantu Divisi Online dalam pengelompokkan pelanggan bedasarkan transaksi mereka. Penambahan fitur ini diharapkan tidak ditemukan lagi kesulitan dalam mengelompokkan data pelanggan mereka. Jadi pembangunan aplikasi *data mining* ini telah sesuai dengan prosedur dan tujuan yang diharapkan.

6. DAFTAR PUSTAKA

[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.