PENERAPAN DATA *MINING* PADA PENJUALAN PRODUK DI DISTRO MATERNAL DISASTER MENGGUNAKAN METODE *CLUSTERING*

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

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

ABSTRAK

Distro Maternal Disaster merupakan tempat penjualan pakaian yang bertempat di Bandung sejak 2003. Produk utama yang dijual di Distro Maternal Disaster adalah pakaian. Setelah melakukan wawancara, didapatkan informasi pihak distro selalu membuat produk baru setiap bulannya dan dikirimkan ke outlet yang telah tersebar. Berdasarkan survey kuesioner terhadap 14 outlet, didapatkan informasi bahwa 11 dari 14 outlet menyatakan produk yang dikirimkan ke setiap outlet tidak sesuai dengan karakteristik outlet. Dilihat dari hasil survey yang telah dilakukan, dapat diambil kesimpulan bahwa produk yang dikirimkan tidak sesuai dengan karakteristik outlet yang berbeda-beda sehingga outlet mengalami kesulitan dalam menjual produk Maternal.

Untuk mendapatkan informasi karakteristik setiap outlet, maka dilakukan segmentasi terhadap setiap outlet menggunakan data *mining* dan metode *clustering*. Algoritma yang digunakan adalah *Agglomerative Hierachical Clustering* (AHC) dimana segmentasi diawali dengan mempresentasikan setiap objek data sebagai satu segmen, kemudian dilakukan perhitungan jarak antar segmen. Setelah itu, segmen yang memiliki jarak terkedat digabungkan menjadi segmen baru. Solusi dari permasalahan tersebut adalah dengan membentuk segmentasi outlet berdasarkan data penjualan, sehingga karakteristik dari setiap outlet dapat terlihat.

Berdasarkan hasil pengujian dapat disimpulkan bahwa segmentasi outlet ini dapat memudahkan pihak Distro untuk membentuk segmentasi outlet sehingga memudahkan pihak distro dalam menentukan strategi bisnis yang tepat untuk setiap segmen outlet berdasarkan karakteristiknya.

Kata kunci : segmentasi outlet, data *mining*, *clustering*, *Agglomeration Hierarchical Clustering*, *complete linkage*

1. PENDAHULUAN

Distro Maternal Disaster merupakan tempat penjualan pakaian yang bertempat di Bandung sejak 2003. Produk utama yang dijual di Distro Maternal Disaster adalah pakaian. Selain pakaian, produk pendukung lainnya seperti tas, dompet, jaket, celana, dan topi yang dipasarkan di beberapa outlet yang tersebar.

Setelah melakukan wawancara dengan pihak distro Maternal Disaster, didapatkan informasi bahwa pihak distro selalu membuat produk baru setiap bulannya dan produk tersebut dikirimkan ke outlet yang telah tersebar. Berdasarkan survey kuesioner terhadap 14 outlet Maternal didapatkan informasi bahwa 11 dari 14 outlet menyatakan produk yang dikirimkan kesetiap outlet tidak sesuai dengan kebutuhan outlet. Ketidaksesusaian tersebut diantaranya karena jenis produk dan ukuran yang tidak sesuai dari setiap outlet, sehingga menjadikan setiap outlet memiliki kebutuhan yang berbeda terhadap produk. Dilihat dari hasil survey yang telah dilakukan, dapat diambil kesimpulan bahwa produk yang dikirimkan tidak sesuai dengan karakteristik outlet yang berbeda-beda sehingga outlet mengalami kesulitan dalam menjual produk Maternal.

Untuk mendapatkan informasi dari setiap outlet, maka diperlukan data penjualan dari setiap outlet pada bulan Agustus 2016 untuk dianalisis. Metode yang dapat digunakan dalam menganalisis banyak adalah metode *data mining*, yaitu metode yang digunakan untuk melihat pola atau set data yang berukuran besar[1]. Terdapat beberapa metode yang dapat digunakan tergantung dari tujuan yang akan dicapai. Salah satu metode yang dapat digunakan adalah metode *clustering*, yaitu metode yang memisahkan data kedalam sejumlah kelompok (*cluster*) menurut kemiripan karakteristik dari data tersebut.

Berdasarkan masalah yang telah dipaparkan sebelumnya, solusi yang mungkin dilakukan adalah dengan membentuk segmentasi outlet sehingga karakteristik kebutuhan produk dari setiap outlet dapat terlihat. Karakteristik outlet tersebut dapat terlihat berdasarkan data penjualan yang telah dilakukan. Maka dari itu, segmentasi akan dilakukan berdasarkan data penjualan. Segmentasi outlet dapat menjadi pengetahuan dengan menganalisis karakteristik atau perilaku berbeda yang membutuhkan produk berbeda dari setiap segmentasinya. Sehingga diharapkan dapat

memaksimalkan proses pengiriman produk sesuai dengan segmentasi outlet yang telah dilakukan.

Maka dari itu Maternal Disaster perlu menggunakan metode *data mining* dan *clustering* dalam menentuka segmentasi outlet dari data penjualan produk yang ada. Oleh karena itu dari penjelasan yang sudah dipaparkan diatas maka perlu dibuat suatu aplikasi "**Penerapan Data Mining pada Penjualan Produk di Distro Maternal Disaster Menggunakan Metode Clustering**".

1.1 Data *Mining*

Data mining adalah kegiatan mengekstraksi atau menambang pengetahuan dari data yang berukuran atau berjumlah besar. Nama lain yang sering digunakan dan memiliki makna yang sama dengan data mining adalah knowledge mining from data, knowledge extraction, pattern analysis, data archaeology, data dredging.

Data mining berusaha memecahkan masalah dengan menganalisis sekumpulan data yang sudah ada dalam basis data. Sedangkan hasil dari proses dari data mining adalah penemuan pengetahuan, informasi, pola, atau model[6].

1.2 *Clustering*

Clustering adalah proses mengelompokkan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas atau *cluster* [7]. Tujuannya menemukan *cluster* yang berkualitas dalam waktu yang layak. *Clustering* dalam data *mining* berguna untuk menemukan pola distribusi di dalam sebuah data set yang berguna untuk proses analisa data. Kesamaan objek biasanya diperoleh dari kedekatan nilai-nilai atribut yang menjelaskan objek-objek data, sedangkan objek-objek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multidimensi.

Dengan menggunakan *clustering*, dapat diidentifikasi daerah yang padat, pola-pola distribusi secara keseluruhan dan keterkaitan yang menarik antara atribut-atribut data. Dalam data *mining* usaha difokuskan pada metode-metode penemuan untuk *cluster* pada basis data berukuran besar secara efektif dan efisien.

Secara garis besar ada beberapa kategori algoritma *clustering* yang dikenal yaitu:

a. Metode Partisi (*Partition-based Clustering*)
   Dalam metode ini pengguna harus menentukan jumlah k partisi yang diinginkan lalu setiap data dites untuk dimasukkan pada salah satu partisi sehingga tidak ada data yang *overlap* dan satu data hanya memiliki satu *cluster*. Contohnya: algoritma *K-Means*, *mixture modeling*.
b. Metode Hierarki (*Hierarchical Clustering*)
   Membuat hierarki berupa dendogram di mana data yang mirip akan disimpan di hierarki yang berdekatan. Metode ini terbagi menjadi dua yaitu *buttom-up* yang menggabungkan *cluster* kecil menjadi *cluster* lebih besar dan *top-down* yang memecah *cluster* besar menjadi *cluster* yang lebih kecil. Contohnya: *Agglomerative Hierarchical Clustering* (AHC), *Divisive Hierarchical Clustering* (DHC).

Dalam metode *clustering* hal yang penting adalah mengukur kemiripan atau ketidakmiripan data yang akan dikelompokkan. *Similaritas* (kemiripan) antara dua objek merupakan ukuran numerik dari seberapa mirip dua buah objek. *Similaritas* mempunyai nilai yang lebih tinggi jika objek-objek tersebut sama. *Range* nilai *similaritas* adalah [0,1]. *Dissimilaritas* (ketidakmiripan) antara dua objek merupakan ukuran numerik dari seberapa berbeda dua buah objek. *Dissimilaritas* bernilai lebih rendah jika objek-objek tersebut mirip. Minimum *dissimilaritas* adalah 0 dan batas atasnya bervariasi. *Dissimiliratas* sering disebut juga dengan *distance* (jarak) [8]. Salah satu cara pengukuran jarak adalah menggunakan jarak *Euclidian* (*Euclidian Distance*), yang dinyatakan dengan persamaan (2.1) berikut :

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

Di mana :
d = jarak
r = jumlah fitur dalam vektor
xi = nilai fitur x ke i
yi = nilai fitur y ke i

1.3 *Agglomeration Hierachical Clustering* (AHC)

Pengelompokan data menggunakan *agglomerative hierarchical clustering* (AHC) diawali dengan merepresentasikan setiap objek pada data sebagai satu kelompok, kemudian dilakukan perhitungan jarak (*distance measure*) antar kelompok tersebut. Setelah itu dua kelompok yang memiliki jarak terdekat digabungkan menjadi sebuah kelompok baru. Kemudian jarak antara kelompok yang baru dengan kelompok lain dihitung menggunakan salah satu metode perhitungan jarak antar kelompok (*single linkage*, *complete linkage*, *average linkage*, dll.) [9]. Proses perhitungan jarak dan penggabungan dua kelompok dilakukan secara iteratif. Berikut ini merupakan beberapa metode yang dapat digunakan untuk menghitung jarak antar kelompok :

a. *Single Linkage* (Jarak Terdekat)
   Single linkage disebut juga *minimum link*, di mana similaritas dari dua *cluster* didasarkan pada dua titik paling dekat dari dua *cluster* yang berbeda. Untuk mencari *single linkage* dapat menggunakan persamaan (2.2) berikut ini :

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

b. *Complete Linkage* (Jarak Terjauh)
   Complete linkage disebut juga *maximum link*, di mana jarak antara dua buah *cluster* dihitung

dari jarak terjauh antara anggota *cluster* satu dengan *cluster* kedua. Untuk mencari *complete linkage* dapat menggunakan persamaan (2.3) berikut ini :

duv=max$\{duv\}$,$duv \in D$ Persamaan (2.3)

c. *Average Linkage* (Jarak Rata-rata)

Average linkage disebut juga *group average*, di mana jarak antara dua buah *cluster* dihitung dari rata-rata jarak antara anggota *cluster* satu dengan *cluster* yang kedua. Untuk mencari *average linkage* dapat menggunakan persamaan (2.4) berikut ini :

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

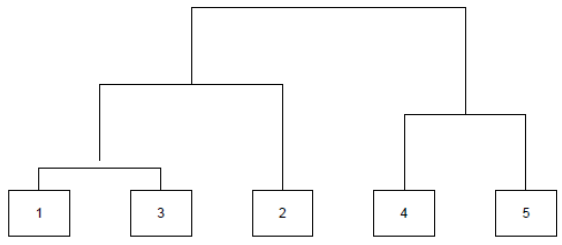
Berikut ini merupakan langkah-langkah dalam algoritma *Agglomerative Hierarchical Clustering* (AHC) [10]:

1. Mempresentasikan setiap outlet sebagai satu kelompok, kemudian menghitung jarak (*distance measure*) antara kelompok-kelompok tersebut.
2. Cari dua *cluster* yang mempunyai jarak paling dekat dengan menggunakan metode *complite linkage* (terdapat pada persamaan (2.3).
3. Ulangi langkah kesatu dan kedua sampai tersisa satu kelompok (kumpulan semua outlet) dan berhenti ketika jumlah *cluster* yang terbentuk sama dengan 1.
4. Ketika jumlah *cluster* berjumlah 1, maka tahap ke-3 telah selesai.

Pemotongan dendrogram dapat dilakukan dengan berdasarkan pada *agglomeration schedule*. Jumlah *cluster* bisa didapatkan dengan mencari perbedaan koofisien jarak dari setiap iterasinya.

1.4 *Dendogram*

Dendogram atau struktur pohon biasanya digunakan untuk merepresentasikan proses *clustering* berbasis hirarki, yang menunjukan bagaimana suatu objek dikelompokan dalam setiap langkahnya. Berikut ini merupakan contoh dari *dendogram* dapat dilihat pada Gambar 1 :



Gambar 1. Dendogram Hasil *Clustering* Berbasis Hirarki

1.5 *Agglomeratin Schedule*

Agglomeration schedule adalah ringkasan numerik dari solusi *cluster* yang dihasilkan dari proses *clustering* menggunakan metode hierarki. [11] Berikut ini merupakan contoh dari *agglomeration schedule* :

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

Gambar 2. Contoh *Agglomeration Schedule*

Penjelasan tabel *agglomeration schedule* :

1. *Stage* merupakan tahapan atau iterasi yang telah dilakukan pada proses *clustering* menggunakan metode hierarki.
2. *Cluster combined* merupakan *cluster* yang digabungkan pada setiap *stage*nya.
3. *Cooficients* merupakan jarak dari dua *cluster* yang digabungkan pada setiap *stagenya* yang dicari dengan menggunakan rumus jarak.
4. *Stage cluster first appear* merupakan *stage* apabila *cluster* yang digabungkan telah digabungkan dengan *cluster* lain pada *stage* sebelumnya. Jika *cluster* yang digabungkan belum pernah digabungkan pada *stage* sebelumnya, maka akan diisi dengan nilai 0.
5. *Next stage* merupakan *stage* selanjutnya saat *cluster* akan digabungkan dengan *cluster* lain. [11].

2. ISI PENELITIAN

Pada bagian ini akan dibahas mengenai tahapan penelitian yang dilakukan.

2.1 Pemahaman Bisnis

Tujuan bisnis dari Distro Maternal Disaster adalah menentukan strategi bisnis yang tepat untuk setiap outlet. Faktor penting agar tujuan bsinis distro Maternal Disaster tercapai yaitu dengan mengetahui karakteristik dari setiap outlet, sehingga produk yang dikirimkan akan terjual sesuai dengan karakteristik outlet tersebut.

Penentuan sasaran data *mining* terdiri dari tujuan dan kriteria sukses data *mining*. Berikut ini merupakan tujuan dan kriteria sukses data *mining* :

1. Tujuan Data *Mining*

   Tujuan dari penerapan data *mining* dalam penelitian ini yaitu untuk membentuk segmentasi outlet di Distro Maternal Disaster. Segmentasi ini akan digunakan oleh bagian *promotion* Distro Maternal Disaster sebagai pertimbangan untuk menentukan strategi bisnis yang tepat untuk setiap outlet.
2. Kriteria Sukses Data *Mining*

   Kriteria sukses dalam penelitian ini adalah mampu membantu bagian *promotion* Ditro Maternal Disaster dalam menentukan segmentasi

outlet yang memiliki karakteristik yang sama dalam penjualan produk.

2.2 Pemahaman Data

Tahap pemahaman data merupakan tahapn kedua yang didalamnya terdapat beberapa tugas, diantaranya :

2.2.1 Pengumpulan Data Awal

Data yang digunakan dalam pengumpulan data awal adalah data penjualan pada bulan Agustus 2016 di distro Maternal Disaster yang berasal dari Microsoft Excel (.xlsx) yang nantinya akan dilakukan proses *mining* pada sebuah aplikasi. Berikut ini merupakan tabel contoh data penjualan pada bulan Agustus 2016 :

MATERNAL DISASTER — LAPORAN DATA PENJUALAN

Arena Experience - Bandung

No	ID Barang	Nama Artikel	Ukuran	Jenis	Sub Jenis	Warna	Harga	Jumlah	Dibayar
1	02607050101	SCR.04	AS	ACCESSORIES	bandana	black	Rp 70.000	1	Rp 70.000
2	02607081001	SCR.05	AS	ACCESSORIES	bandana	red	Rp 70.000	1	Rp 70.000
3	03207041201	BHOF	AS	ACCESSORIES	belt	brown	Rp 295.000	3	Rp 885.000
4	03407100102	FG.01	AS	ACCESSORIES	gloves	black	Rp 130.000	1	Rp 130.000
5	03407100101	GLV.01	AS	ACCESSORIES	gloves	black	Rp 160.000	1	Rp 160.000
6	02707140101	FCCKS	AS	ACCESSORIES	socks	black	Rp 60.000	2	Rp 120.000
7	02707140501	BOLOCK WHITE	AS	ACCESSORIES	socks	white	Rp 60.000	2	Rp 120.000
8	02907140102	THSOCK	AS	ACCESSORIES	socks	black	Rp 60.000	1	Rp 60.000
9	03407080101	LODGE	AS	ACCESSORIES	wallets	black	Rp 330.000	2	Rp 660.000
10	03303011701	GALAD	AS	BAG	backpack	olive green	Rp 310.000	1	Rp 310.000
11	02903011601	GOTHA	AS	BAG	backpack	dark green	Rp 290.000	4	Rp 1.160.000
12	02703020101	RAPHS	AS	BAG	sling bag	black	Rp 275.000	4	Rp 1.100.000
13	02703021601	DAGTF	AS	BAG	sling bag	dark green	Rp 275.000	3	Rp 825.000
14	02903020101	NILECA	AS	BAG	sling bag	black	Rp 290.000	2	Rp 580.000
15	02903020102	CALLER	AS	BAG	sling bag	black	Rp 290.000	1	Rp 290.000
16	03203060801	GLUTO	AS	BAG	waist bag	navy blue	Rp 225.000	2	Rp 450.000
17	02603060301	NAIWE	AS	BAG	waist bag	gray	Rp 205.000	1	Rp 205.000
18	02603060102	GATHER	AS	BAG	waist bag	black	Rp 205.000	1	Rp 205.000
19	02703060103	SCTV	AS	BAG	waist bag	black	Rp 210.000	1	Rp 210.000

Gambar 3. Data Penjualan Bulan Agustus 2016

2.2.2 Penjelasan Data

Data awal yang telah dikumpulkan akan dijelaskan pada tahap ini. Pada data terdapat 10 atribut, yaitu No, ID Barang, Nama Artikel, Ukuran, Jenis, Sub Jenis, Warna, Harga, Jumlah, Dibayar.

2.2.3 Eksplorasi Data

Tahapan eksplorasi data dapat membantu tujuan *data mining.* Pada tahap ini, dilakukan analisis secara mendalam terhadap data yang telah diperoleh. Kegiatan ini dilakukan untuk mengetahui karakteristik dan kualitas data tersebut. Namun pada tahap ini eksplorasi data tidak dilakukan karena data yang disajikan sudah jelas tanpa harus melakukan tahap eksplorasi data.

2.2.4 Evaluasi Kualitas Data

Evaluasi kualitas data dilakukan terhadap atribut jumlah penjualan berdasarkan jenis, sub jenis, warna, dan ukuran. Berikut ini merupakan hasil evaluasi dari atribut-atribut tersebut.

1. Tahap eksplorasi tidak diperlukan pada kasus ini karena data yang disajikan sudah jelas tanpa harus melakukan tahap eksplorasi data. Eksplorasi data dilakukan untuk mengetahui kualitas data.
2. *Missing value* mungkin saja ditemukan pada data penjualan. Akan tetapi pada data penjualan *missing value* tidak ditemukan.

2.3 Persiapan Data

Pada tahap ini akan dilakukan pengolahan terhadap data yang telah diperoleh (data *preprocessing*). *Data preprocessing* perlu dilakukan untuk meningkatkan kualitas data agar dapat digunakan pada proses pemodelan, terdapat beberapa tugas dalam tahap ini, yaitu:

2.3.1 Pemilihan Data

Pemilihan data dilakukan untuk menentukan data yang akan digunakan pada tahap pemodelan atau proses *clustering*. Pemilihan data tersebut meliputi pemilihan atribut dan *record* yang akan ada dalam sistem. terdapat 4 atribut yang akan dipilih untuk tahap pemodelan yaitu jenis, sub jenis, warna, ukuran, dan jumlah. Atribut dan *record* tersebut dipilih karena sesuai dengan tujuan bisnis dari distro Maternal Disaster.

2.3.2 Pembersihan Data

Proses pembersihan data dilakukan untuk meningkatkan kualitas data yang diperoleh. Pembersihan data ini mengacu pada evaluasi kualitas data yang sebelumnya telah dilakukan. Jika terdapat *missing value*, maka nilai yang hilang tersebut akan diisi dengan nilai *default* 0. Tidak dilakukan analisis terhadap *outlier* karena algoritma yang digunakan tidak terpengaruh oleh nilai *outlier*.

2.3.3 Penyusunan Data

Pada proses ini dilakukan pembangunan data set baru atau memproduksi atribut turunan melalui transformasi data yang sesuai dengan kebutuhan pemodelan. Transformasi pada analisis kali ini melibatkan data penjualan pada bulan Agustus 2016, menurunkan atribut-atribut yang diperlukan pada tahap pemodelan dan membuang atribut yang tidak diperlukan untuk tahap pemodelan. Atribut yang dipilih pada penyusunan data ini adalah jenis, sub jenis, warna, dan ukuran yang digabungkan menggunakan tanda *underscore* dan diberi nama Data transaksi penjualan. Berikut hasil data transaksi penjualan adalah sebagai berikut :

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

Gambar 4. Penyusunan Data

2.3.4 Integrasi Data

Pada tahap ini dilakukan penggabungan dua tabel atau lebih yang memiliki informasi berbeda tentang objek yang sama kedalam set data baru. Namun pada tahap ini, integritas tidak dilakukan, karena data yang diperlukan tidak perlu melakukan penggabungan.

2.4 Pemodelan

Tahap pemodelan merupakan kegiatan utama dari proses penyelesaian masalah data *mining*. Data yang sudah disusun akan diproses menggunakan suatu metode sehingga dapat menghasilkan model. Kegiatan pada tahap pemodelan meliputi teknik pemodelan yang digunakan, mekanisme dari model yang digunakan, serta analisis pengujian model.

2.4.1 Teknik Pemodelan

Teknik pemodelan yang akan digunakan yaitu menggunakan *herarchical clustering*, dan model yang akan digunakan adalah algoritma *Agglomerative Herarchical Clustering* (AHC). Metode tersebut digunakan karena dapat membentuk segmentasi outlet secara alami berdasarkan kemiripan nilai-nilai atribut pada data. Selain itu, jumlah segmentasi outlet yang akan dibentuk belum diketahui. Dengan menggunakan metode tersebut, jumlah segmentasi yang diinginkan dapat ditentukan setelah model terbentuk. Hasil segmentasi outlet akan disajikan pada sebuah *dendogram*.

2.4.2 Mekanisme Pemodelan

Pada tahap ini dijelaskan mekanisme dari model yang akan digunakan serta asumsi awal dari model. Model algoritma *agglomerative hierarchical clustering* memiliki tahapan sebagai berikut :

5. Mempresentasikan setiap outlet sebagai satu kelompok, kemudian menghitung jarak (*distance measure*) antara kelompok-kelompok tersebut.
6. Cari dua *cluster* yang mempunyai jarak paling dekat dengan menggunakan metode *complite linkage* (terdapat pada persamaan (2.3).
7. Ulangi langkah kesatu dan kedua sampai tersisa satu kelompok (kumpulan semua outlet) dan berhenti ketika jumlah *cluster* yang terbentuk sama dengan 1.
8. Ketika jumlah *cluster* berjumlah 1, maka tahap ke-3 telah selesai.

2.4.3 Penggunaan Teknik Pemodelan

Pada tahap ini, peneliti akan menerapkan metode pemodelan yang telah dijelaskan sebelumnya. Berikut merupakan proses penerapan metode *Hierarchical Agglomerative Clustering* pada data yang telah disiapkan.

Langkah ke-1

Inisialisasi *cluster* awal adalah sebanyak 14 *cluster*, karena data(n) = *cluster*(n), kemudian hitung jarak menggunakan *Euclidean Distance* dihitung menggunakan persamaan berikut :

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

Setiap *cluster* akan dihitung jarak dirinya sendiri dengan *cluster* lainnya. Berikut ini merupakan proses perhitungan jarak antar *cluster*. Berikut ini merupakan contoh perhitungan jarak berdasarkan ukuran :

Dman(1,2) =

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

= 40,78

Perhitungan dengan menggunakan persamaan *EuclideanDistance* dilakukan terhadap seluruh data, sehingga akan dihasilkan matriks jarak. Untuk lebih lengkap matriks jarak dapat dilihat pada Lampiran F.

Langkah ke-2

Pada tahapan ini akan dihitung jarak antar dua *cluster* dengan menggunakan *complete linkage* (persamaan 2.3). dengan menggunakan metode ini akan dicari jarak yang paling dekat dari dua *cluster* dan menggabung *cluster* dengan jarak maksimum dan menghasilkan *cluster* baru. Berikut contoh perhitungan dari matriks *Euclidean Distance*.

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

Gambar 5. Matriks *Euclidean Distance*

Min(Dman) = min (D(3,13) = 19,03. Dari semua *cluster* yang ada, terpilih *cluster* 3 dan 13 karena memiliki nilai paling minimum dari *cluster* lain.

Menghitung jarak 3 dan 13 dengan *cluster* lain yang tersisa. Dengan contoh sebagai berikut :

D(3,13) 1 = maks{d(3,1) d(13,1)} = maks{(36,10) (34,77)} = 36,10. Setelah membandingan dua nilai (36,10) dan (34,77) dipilih nilai maksimum. Setelah dilakukan perhitungan jarak terhadap semua *cluster*, d(3,13) akan membentuk *cluster* baru, sehingga *cluster* 3 dan 13 akan dihapus. Dengan contoh sebagai berikut:

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

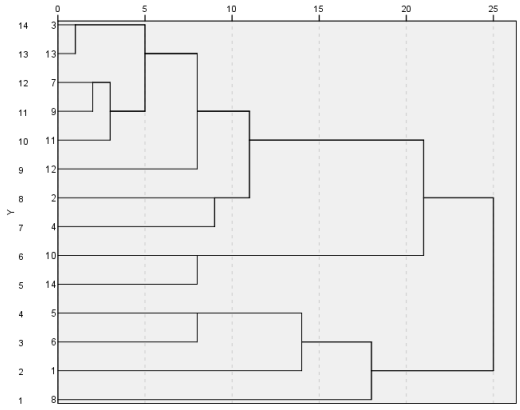
Gambar 6. Matriks *Euclidean Distance* Iterasi ke-1

Langkah ke-3

Ulangi langkah kesatu dan kedua sampai tersisa satu kelompok (kumpulan semua outlet) dan akan berhenti ketika jumlah *cluster* yang terbentuk sama dengan 1.

Langkah ke-4

Ketika jumlah *cluster* berjumlah 1,maka tahap 3 telah selesai dan hasil akhir dari perulangan berdasarkan ukuran dapat dilihat pada Lampiran D. berikut ini merupakan *dendogram* dari hasil pengelompokkan.seperti terlihat pada Gambar 7. Hasil Akhir *Dendogram.*



Gambar 7. Hasil Akhir *Dendogram*

Berdasarkan *dendogram* yang telah dibentuk, tingkat yang tepat atau jumlah *cluster* terbaik akan ditentukan dengan melihat selisih jarak dari setiap *cluster*yang terbentuk. Untuk mengetahui jumlah *cluster* terbaik, salah satunya dengan menggunakan *agglomeration schedule. Agglomeration schedule* dapat dilihat pada Gambar 8.

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

Gambar 8. *Agglomeration Schedule*

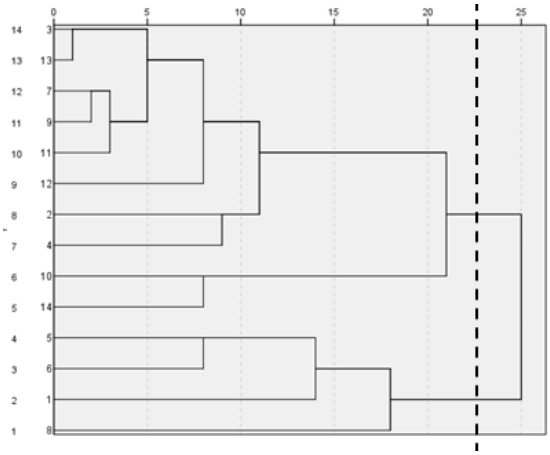
Pada tahap pertama, *cluster* yang digabungkan adalah *cluster* 3 dan *cluster* 13, karena memiliki jarak paling kecil yaitu 19,03. *Cluster* yang telah digabungkan pada iterasi kedua akan digabungkan kembali pada iterasi ketiga dan iterasi keempat akan digabungkan kembali pada iterasi kelima. Pada iterasi ke-9, cluster yang telah digabungkan pada *cluster* 5 dan *cluster* 8 akan digabungkan kembali. *Cluster* yang akan dihasilkan berikutnya akan muncul pada iterasi ke-12, begitu seterusnya hingga iterasi 13.

Jumlah *cluster* yang terbaik dapat dilihat dari selisih koefisien dari setiap iterasinya. Selisih dari koefisien segmentasi berdasarkan ukuran dapat dilihat pada Gambar 9.

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

Gambar 9. Selisih Koefisien Pengelompokan

Selisih terbesar yaitu 4,37 sehingga jumlah *cluster* terbaik adalah 2 *cluster*. Maka *dendogram* hasil pengelompokkan berdasarkan ukuran akan dilakukan pemotongan seperti terlihat pada Gambar 10.



Gambar 10. Pemotongan *Dendogram*

Hasil akhir dari segmentasi ini, terbentuk 2 *cluster* dengan masing-masing kelompok dapat dilihat pada Tabel 1 dan Tabel 2.

Tabel 1. *Cluster* 1

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

Tabel 2. *Cluster* 2

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

Dari hasil segmentasi outlet yang telah dilakukan berdasarkan penjualan produk, didapatkan hasil penjualan tertinggi dari setiap *cluster* adalah sebagai berikut :

Tabel 3. Penjualan Tertinggi

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

Dari hasil penjualan tertinggi tersebut, dapat dihasilkan rekomendasi produk untuk setiap *cluster* yang dapat dilihat pada Tabel 4 berikut :

Tabel 4. Rekomendasi Produk Setiap *Cluster*

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

3. PENUTUP

Hasil dari penelitian yang telah dilakukan, diperoleh kesimpulan bahwa aplikasi pembentukan ssegmentasi outlet yang telah dibangun dapat membantu manager *promotion* untuk membentuk segmentasi outlet yang sesuai dengan kemiripan penjualan produk sehingga dapat dijadikan acuan untuk pendistribusian dan strategi bisnis yang tepat untuk karakteristik setiap outlet nya.

Ada beberapa saran yang dapat dilakukan untuk pengembangan aplikasi pembentukan segmentasi outlet ini, antara lain :

1. Menambah atribut yang digunakan dalam pemodelan, agar hasil segmentasi lebih maksimal dalam menentukan karakteristik outlet.
2. Algoritma yang digunakan dapat dikembangkan lagi dengan menggunakan algoritma lain, sehingga dapat membandingkan hasil segmentasinya dengan algoritma *complete linkage*.
3. Metode yang digunakan dalam perhitungan jarak bisa dikembangkan lagi dengan menggunakan metode *Manhattan*, karena perhitungan dengan metode ini lebih mudah dipahami jika dibandingkan dengan metode pencarian jarak *euclidean.*

DAFTAR PUSTAKA

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

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

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

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

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

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

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

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

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

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

[11] M. Caccam dan J. Refran, "Cluster Analysis".
[12] A.S. Rosa dan M. Shalahuddin, Modul Pembelajaran Rekayasa Perangkat Lunak (Terstruktur dan Berorientasi Objek), Bandung: Modula, 2013.
[13] Ir. Supardi, Y., "Semua Bisa Menjadi Programming Java Basic Programming", Jakarta: Elex Media,2010
[14] Simarmata, Janner., "*Rekayasa Perangkat Lunak",* Yogyakarta : Andi, 2010