

PENERAPAN DATA MINING PADA DATA PENJUALAN SEPATU UNTUK MEMBENTUK SEGMENTASI DISTRIBUTOR DI CV. XYZ MENGGUNAKAN METODE CLUSTERING

Ririn Patmawati Sari¹, Adam Mukharil Bachtiar²

^{1,2} Universitas Komputer Indonesia

Jalan Dipatiukur No. 112-116, Bandung 40132, Indonesia

E-mail : rfatmawatisari@gmail.com¹, adam@email.unikom.ac.id²

ABSTRAK

CV. XYZ merupakan produsen sepatu bermotif etnik yang memasarkan produknya pada distributor-distributor yang tersebar di beberapa wilayah di Indonesia. Pendistribusian sepatu saat ini kurang maksimal karena tidak sesuai dengan kebutuhan produk di distributor. Dampaknya yaitu banyaknya distributor yang menukarkan sepatu dan proses distribusi untuk sampai ke tangan konsumen terkadang terhambat. Solusi dari permasalahan tersebut adalah dengan membentuk segmentasi distributor berdasarkan data penjualan. Segmentasi distributor tersebut dapat menjadi pengetahuan dengan menganalisis karakteristik atau perilaku yang berbeda yang mungkin membutuhkan produk atau pemasaran yang berbeda dari setiap segmennya. Pembentukan segmentasi distributor dapat dilakukan dengan memanfaatkan data *mining* metode *clustering* dengan menggunakan algoritma AHC. Algoritma AHC membentuk *cluster* berdasarkan kedekatan jarak di setiap *clusternya*. Berdasarkan hasil pengujian dapat disimpulkan bahwa penelitian ini dapat membantu pihak pemasaran CV. XYZ untuk melakukan pemasaran dan strategi bisnis yang tepat untuk setiap segmen distributor sesuai dengan karakteristik di setiap segmennya.

Kata Kunci : segmentasi distributor, data *mining*, *clustering*, *agglomerative hierarchical clustering*, AHC

1. PENDAHULUAN

CV. XYZ merupakan produsen sepatu bermotif etnik yang memasarkan produknya pada distributor-distributor yang tersebar di beberapa wilayah di Indonesia. Distributor-distributor tersebut yang melayani pelanggan secara langsung, sehingga keberadaan dari distributor-distributor ini memiliki peran yang sangat penting untuk menentukan strategi bisnis ke depannya.

Setelah melakukan wawancara dengan pihak CV. XYZ, didapatkan informasi bahwa proses

pendistribusian produk akan dilakukan setelah distributor meminta kepada CV. XYZ, distributor meminta produk apabila persediaan produk di distributor telah habis atau hanya tersisa sedikit. Setelah melakukan permintaan kepada pihak CV. XYZ, sepatu akan didistribusikan ke distributor dan sepatu yang didistribusikan tersebut tergantung persediaan yang ada di gudang. Proses pendistribusian tersebut dirasa kurang maksimal. Hal tersebut dikarenakan pendistribusian sepatu tidak sesuai dengan kebutuhan produk di distributor. Dampaknya yaitu banyaknya distributor yang menukarkan sepatu dan proses distribusi untuk sampai ke tangan konsumen terkadang terhambat.

Berdasarkan masalah yang telah dipaparkan sebelumnya, solusi yang mungkin dilakukan adalah dengan membentuk segmentasi distributor sehingga karakteristik kebutuhan produk dari setiap segmennya dapat terlihat. Kebutuhan dari setiap kelompoknya dapat terlihat berdasarkan data penjualan yang telah dilakukan. Maka dari itu segmentasi akan dilakukan berdasarkan data penjualan. Segmentasi distributor tersebut dapat menjadi pengetahuan dengan menganalisis karakteristik atau perilaku yang berbeda yang mungkin membutuhkan produk atau pemasaran yang berbeda dari setiap segmennya. Sehingga diharapkan dapat memaksimalkan proses pemasaran dan pendistribusian sepatu di CV. XYZ.

Dikarenakan jumlah distributor dan data penjualan yang cukup banyak mengakibatkan diperlukannya metode yang tepat untuk membentuk segmentasi distributor. Metode yang dapat digunakan dalam menganalisis data yang sangat banyak adalah metode data *mining*, yaitu metode yang digunakan untuk melihat pola atau set data yang berukuran besar [1]. Dalam data *mining* 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 ke dalam sejumlah kelompok (*cluster*) menurut kemiripan karakteristik dari data tersebut.

Maka dari itu CV. XYZ memerlukan penerapan data *mining* dengan metode *clustering* untuk membentuk segmentasi distributor berdasarkan data penjualan.

2. ISI PENELITIAN

Pada bagian ini akan dibahas mengenai tahapan penelitian yang dilakukan.

2.1 Pemahaman Bisnis

Tujuan bisnis dari CV. XYZ adalah menjadi *brand* sepatu *handmade* terbesar di Indonesia. Faktor penting agar tujuan bisnis CV. XYZ tercapai yaitu dengan menjaga hubungan baik dengan distributor, dan terus meningkatkan pelayanan terhadap distributor dalam penyediaan dan pendistribusian produk yang sesuai dengan permintaan pasar.

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 kelompok distributor di CV. XYZ. Kelompok ini akan digunakan oleh bagian pemasaran CV. XYZ sebagai pertimbangan strategi pendistribusian produk kepada distributor.

2. Kriteria sukses data *mining*

Kriteria sukses dalam penelitian ini yaitu apabila mampu membantu bagian pemasaran CV. XYZ dalam mengelompokkan distributor yang memiliki karakteristik yang sama dalam penjualan produk.

2.2 Pemahaman Data

Tahap pemahaman data merupakan tahapan kedua yang di dalamnya terdapat beberapa tugas, yaitu :

2.2.1 Pengumpulan Data Awal

Data awal yang telah dikumpulkan adalah data penjualan dari bulan Januari sampai April 2016. Data yang digunakan untuk pemodelan pada penelitian ini adalah data penjualan pada bulan Januari 2016. Data penjualan yang akan digunakan tersebut berupa *file* yang memiliki format *.xls atau *.xlsx. Berikut ini merupakan sebagian data awal yang merupakan data penjualan pada bulan Januari 2016 :

No	Tanggal	ID Distributor	Distributor	Alamat	TS					Total	Total Harga	
					35	36	37	38	39			40
1	5-Jan-16	1	Rani	Majalengka	0	1	0	0	0	0	26	2.444.000.00
2	8-Jan-16	2	Kurniawan	Bandung Barat	0	0	0	2	0	0	54	5.076.000.00
3	9-Jan-16	3	Wulandari	Medan	0	0	0	1	0	0	24	2.256.000.00
4	13-Jan-16	4	Eni	Surabaya	0	0	1	0	1	0	53	4.982.000.00
5	15-Jan-16	5	Poppy	Bandung	0	0	3	0	0	0	23	2.162.000.00
6	15-Jan-16	6	Renny	Jakarta Barat	0	0	0	0	0	0	47	4.418.000.00
7	16-Jan-16	7	Hani	Bandung	0	5	2	2	0	0	55	5.170.000.00
8	18-Jan-16	8	Endri	Yogyakarta	3	0	0	0	0	0	28	2.632.000.00
9	18-Jan-16	9	Ibeth	Tangerang	0	0	0	0	0	0	21	1.974.000.00
10	21-Jan-16	10	Eerstiana	Tangerang	0	0	0	0	3	0	17	1.598.000.00
11	22-Jan-16	11	Jay Ezah	Brunel Darussalam	0	7	5	4	0	0	58	5.452.000.00
12	23-Jan-16	12	Herpi	Lombok	0	5	2	2	0	0	59	5.546.000.00
13	25-Jan-16	13	Zuhadha	Bogor	0	0	0	0	0	0	16	1.506.000.00
14	25-Jan-16	14	Cut Opie	Lhokseawe	1	8	3	4	0	0	64	6.016.000.00
15	27-Jan-16	15	Ema	Lamongan	0	0	0	0	0	0	22	2.068.000.00
16	27-Jan-16	16	Lina	Sidoarjo	0	0	0	0	0	0	15	1.410.000.00
17	27-Jan-16	17	Misrah	Cimahi	0	0	0	0	0	0	16	1.504.000.00
18	27-Jan-16	18	Dewi	Soreang	0	0	0	1	0	0	23	2.162.000.00
19	28-Jan-16	19	Mike Sariang	Stubondo	0	0	0	2	0	0	50	4.700.000.00
20	29-Jan-16	20	Svanah	Serang	0	0	0	0	0	0	41	3.854.000.00

Gambar 1. Data Penjualan Bulan Januari 2016

2.2.2 Penjelasan Data

Data awal yang telah dikumpulkan akan dijelaskan pada tahap ini. Pada data terdapat 127 atribut, yaitu No, Tanggal, ID distributor, nama distributor, alamat, serta penjualan dari setiap motif dan nomor sepatu. Setiap motif memiliki 10 ukuran sepatu yaitu 35 – 44 dan terdapat 12 motif yaitu TS, BR, PR, SP, RV, GP, QZ, HD, YN, HB, PD dan RN.

2.2.3 Eksplorasi Data

Tahapan eksplorasi data dapat membantu tujuan dari data *mining*. Dalam penelitian ini eksplorasi data meliputi analisis deskriptif. Data yang akan digunakan untuk tahap eksplorasi data ini adalah data penjualan sepatu berdasarkan motif dan ukuran pada data penjualan bulan Januari 2016.

1. Analisis Deskriptif

Analisis deskriptif ini dilakukan untuk mendeskripsikan atau menjelaskan kelompok data. Atribut yang akan digunakan dalam analisis deskriptif ini adalah atribut jumlah penjualan sepatu berdasarkan motif dan ukuran. Analisa dilakukan pada atribut tersebut karena atribut pengelompokan yang akan dilakukan berdasarkan pada jumlah penjualan sepatu berdasarkan motif dan ukuran. Ada beberapa nilai yang akan dicari dalam analisis deskriptif ini, yaitu :

a. Nilai Minimum

Nilai minimum digunakan untuk mengetahui berapa jumlah penjualan yang paling sedikit. Dengan mengetahui nilai minimum akan memberikan gambaran berapa nilai terkecil dari data penjualan produk sebelum dilakukan proses data *mining*.

b. Nilai Maksimum

Nilai maksimum digunakan untuk mengetahui berapa jumlah penjualan yang paling banyak. Dengan mengetahui nilai maksimum akan memberikan gambaran berapa nilai terbesar dari data penjualan produk sebelum dilakukan proses data *mining*.

c. Nilai Rata-rata

Nilai rata-rata digunakan untuk mengetahui berapa jumlah rata-rata penjualan produk. Dengan mengetahui nilai rata-rata ini akan memberikan gambaran berapa nilai rata-rata dari data penjualan produk sebelum dilakukan proses data *mining*. Nilai rata-rata didapatkan dengan menggunakan persamaan berikut :

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \tag{1}$$

d. Standar Deviasi

Standar deviasi ini digunakan untuk menggambarkan bagaimana penyebaran data dari rata-rata jumlah. Dengan mengetahui nilai standar deviasi ini maka akan memberikan bagaimana simpangan rata-rata dari data yang akan diteliti sebelum proses data *mining*.

Standar deviasi dicari menggunakan persamaan berikut ini :

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} \quad (2)$$

e. Nilai Kuartil dan IQR (*Interquartile Range*)

Nilai kuartil ini dapat digunakan untuk mendeteksi nilai *outlier*. Nilai kuartil dan IQR dapat dicari dengan menggunakan persamaan berikut ini :

$$Letak Q_i = \frac{i(n+1)}{4} \quad (3)$$

2.2.4 Evaluasi Kualitas Data

Evaluasi kualitas data dilakukan terhadap atribut jumlah pembelian berdasarkan motif dan ukuran. Berikut ini merupakan hasil evaluasi dari atribut-atribut tersebut :

- a. *Missing value* mungkin saja ditemukan pada data penjualan tetapi pada data penjualan bulan Januari 2016 yang terdapat pada lampiran F, tidak terdapat *missing value*.
- b. Berdasarkan data penjualan bulan Januari 2016 akan dilihat apakah terdapat *outlier* atau tidak, persamaan berikut ini merupakan cara yang dapat dilakukan untuk menemukan data *outlier*:

$$Batas bawah = Q_3 + (3 \times IQR) \quad (4)$$

$$Batas atas = Q_3 + (6 \times IQR) \quad (5)$$

2.3 Persiapan Data

Tahapan ini mencakup semua kegiatan yang diperlukan untuk membangun data set akhir (data yang akan dimasukkan ke dalam pemodelan), terdapat beberapa tugas dalam tahap ini, yaitu:

2.3.1 Pemilihan Data

Pemilihan data memiliki tugas meliputi pemilihan atribut. Atribut yang dipilih untuk melakukan proses pengelompokan (*clustering*). Atribut yang akan dipakai dalam proses *clustering* antara lain, ID distributor dan jumlah penjualan sepatu berdasarkan motif dan ukuran.

2.3.2 Pembersihan Data

Pada proses ini dilakukan proses pembersihan data, berdasarkan hasil evaluasi kualitas diperoleh hasil sebagai berikut:

- 1. Pada data penjualan bulan Januari 2016 *missing value* tidak ditemukan, tetapi jika terdapat *missing value*, maka nilai yang hilang tersebut akan diisi dengan nilai default yaitu 0.
- 2. Nilai yang bersifat *outlier* akan tetap diproses, karena algoritma yang digunakan tidak terlalu terpengaruh oleh *outlier*.

2.3.3 Penyiapan Data Awal

Pada proses penyiapan data awal akan disiapkan data yang akan digunakan di tahap pemodelan. Data yang akan digunakan dalam tahapan pemodelan adalah data penjualan bulan Januari 2016 yang telah mengalami proses perubahan atribut untuk motif dan nomor sepatu yang digabungkan dengan tanda *underscore*.

ID Distributor	TS 35	TS 36	TS 37	TS 38	TS 39	TS 40	RN 42	RN 43	RN 44
1	0	1	0	0	0	0	1	5	2
2	0	0	0	2	0	0	0	3	4
3	0	0	0	1	0	0	0	0	0
4	0	0	1	0	1	0	0	1	2
5	0	0	3	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
7	0	5	2	2	0	0	0	0	0
8	3	0	0	0	0	0	0	4	0
9	0	0	0	0	0	0	0	0	0
10	0	0	0	0	3	0	0	0	0
11	0	7	5	4	0	0	0	0	0
12	0	5	2	2	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0
14	1	8	3	4	0	0	0	3	2
15	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0
18	0	0	0	1	0	0	0	0	0
19	0	0	0	2	0	0	0	3	3
20	0	0	0	0	0	0	0	0	0

Gambar 2. Penyiapan Data Awal

2.4 Pemodelan

Pada tahap ini akan dilakukan pemodelan, pada tahap ini memiliki beberapa tugas yaitu pemilihan teknik pemodelan, penjelasan mekanisme model, serta pengujian model pada data yang telah disiapkan sebelumnya.

2.4.1 Teknik Pemodelan

Teknik pemodelan yang akan digunakan yaitu menggunakan *hierarchical clustering*, dan model yang akan digunakan yaitu algoritma *Agglomerative Hierarchical Clustering* (AHC).

2.4.2 Mekanisme Pemodelan

Tahap ini menjelaskan mengenai mekanisme dari model yang akan digunakan serta asumsi awal dari model. Model algoritma *agglomerative hierarchical clustering* memiliki tahapan sebagai berikut:

- 1. Setiap objek data dianggap sebagai *cluster*. Jika n = jumlah data dan k = jumlah *cluster*, maka k = n.
- 2. Menghitung matriks jarak antar *cluster* (menggunakan *euclidean distance* yang terdapat pada persamaan (2.6)).
- 3. Cari dua *cluster* yang mempunyai jarak paling dekat (*single linkage* yang terdapat pada persamaan (2.7)) lalu gabungkan dua *cluster* tersebut, sehingga k = k-1.
- 4. Jika k > 1, ulangi langkah 2 dan 3.
- 5. Jika k = 1, perulangan selesai.
- 6. Menghasilkan *cluster* dengan memotong *dendrogram* pada tingkat yang tepat.

2.4.3 Analisis Pengujian Model

Data yang digunakan untuk melakukan pemodelan ini adalah data penjualan pada bulan Januari 2016. Jumlah data yang akan digunakan adalah sebanyak 20 record, seperti terlihat pada Tabel 3.12. Kasus yang akan diuji dengan menggunakan algoritma AHC adalah sebagai berikut:

- 1. Ubah setiap data menjadi cluster
- Inisialisasi cluster awal adalah sebanyak 20 cluster karena data (n) = cluster (k), kemudian dihitung jarak antara cluster dengan cluster yang ada dengan menggunakan persamaan Euclidean Distance.

2. Menghitung jarak antar data

Dalam tahap ini setiap cluster akan dihitung jarak dengan dirinya sendiri dan jarak dengan cluster lainnya. *Euclidean distance* dihitung menggunakan persamaan berikut :

$$dist = \sqrt{\sum_{k=1}^n (p_k - q_k)^2} \quad (6)$$

Berikut ini merupakan proses perhitungan jarak antar cluster dengan menggunakan persamaan *Euclidean Distance*. Berikut ini contoh perhitungan jarak berdasarkan ukuran :

$$d(1,1) = \sqrt{\begin{matrix} (0-0)^2 + (1-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ \dots \\ + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \end{matrix}} = 0$$

$$d(1,2) = \sqrt{\begin{matrix} (0-0)^2 + (1-0)^2 + (0-0)^2 + (0-2)^2 \\ + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ + (0-3)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \\ \dots \\ + (0-3)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 \end{matrix}} = 12.49$$

Dengan cara yang sama yaitu dengan menggunakan persamaan *Euclidean Distance* dilakukan perhitungan terhadap seluruh data. Dari proses perhitungan *Euclidean Distance* maka akan dihasilkan matriks jarak.

3. Mencari dua cluster yang saling berdekatan

Pada tahapan ini akan dihitung jarak antara dua cluster dengan menggunakan *single linkage* (persamaan 2.7), dengan metode ini akan mencari jarak yang paling dekat dari dua cluster.

Iterasi ke-1 :

$$d_{(1,2)3} = \min\{d_{(1,3)}, d_{(2,3)}\} = \min\{18.73, 15.49\} = 15.49$$

sampai dengan :

$$d_{(u,v)n} = \min\{d_{(u,n)}, d_{(v,n)}\}$$

Dengan metode *single linkage* pada setiap iterasi dua cluster yang memiliki jarak paling kecil akan digabungkan. Jika *cluster(n,n)* dan *euclidean distance* = 0, maka nilai *euclidean distance* tersebut tidak dinyatakan sebagai yang terkecil dan pencarian jarak terkecil akan dilanjutkan. Misalkan *cluster(1,1)*, *cluster(2,2)*, *cluster(3,3)*, dan seterusnya.

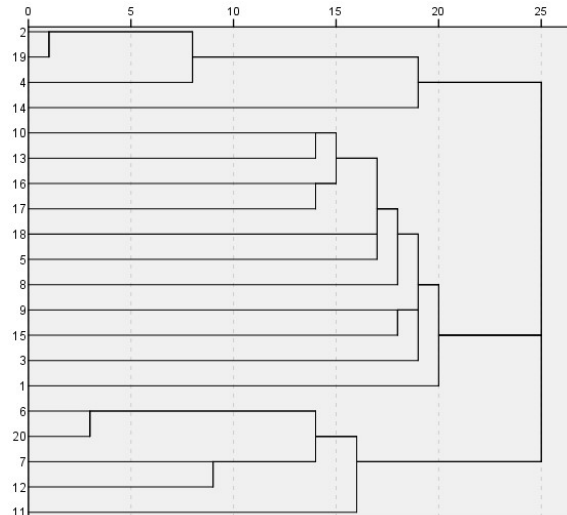
Pada saat pengelompokan, iterasi pertama jarak yang paling dekat adalah cluster 2 dan cluster 19 dengan nilai jarak yaitu 2.00. Maka cluster tersebut akan digabungkan menjadi satu cluster dan jumlah cluster baru yang terbentuk sebanyak 19 cluster.

4. Jika k > 1, ulangi langkah 2 dan 3

Langkah ke-2 dan ke-3 terus dilakukan dan akan berhenti ketika jumlah cluster yang terbentuk sama dengan 1. Untuk mendapatkan 1 cluster, dalam kasus ini perlu dilakukan sebanyak 19 iterasi.

5. Jika k = 1, perulangan selesai

Jumlah cluster akan berjumlah 1 saat perulangan ke-19. Berikut ini merupakan dendrogram dari hasil pengelompokan.



Gambar 3. Hasil Akhir Dendrogram

6. Menghasilkan cluster dengan memotong dendrogram pada tingkat yang tepat

Berdasarkan dendrogram 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 satu caranya dapat menggunakan *agglomeration schedule*.

Iterasi	Kombinasi Cluter		Koofisien	Iterasi Kombinasi Cluster sebelumnya		Iterasi Selanjutnya
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	2	19	2.00	0	0	3
2	6	20	3.16	0	0	5
3	2	4	5.20	1	0	14
4	7	12	5.48	0	0	5
5	6	7	7.21	2	4	9
6	10	13	7.42	0	0	8
7	16	17	7.55	0	0	8
8	10	16	7.81	6	7	10
9	6	11	8.19	5	0	19
10	10	18	8.43	8	0	11
11	5	10	8.72	0	10	13
12	9	15	9.00	0	0	15
13	5	8	9.11	11	0	16
14	2	14	9.38	3	0	18
15	3	9	9.43	0	12	16
16	3	5	9.49	15	13	17
17	1	3	9.95	0	16	18
18	1	2	11.75	17	14	19
19	1	6	12.00	18	9	0

Gambar 4. Agglomeration Schedule

Pada tahap pertama, cluster yang digabungkan adalah cluster 2 dan cluster 19, karena memiliki jarak yang paling kecil yaitu 2.00.

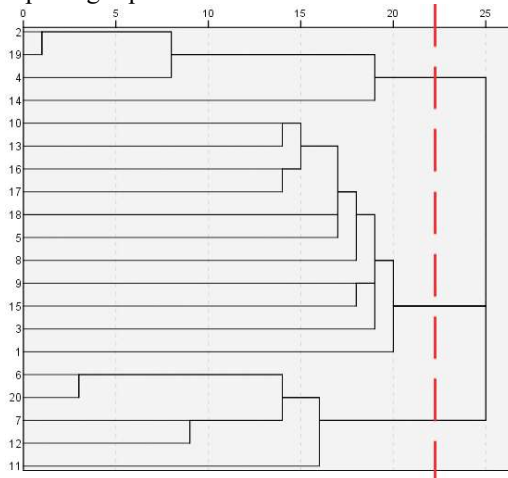
Cluster yang telah digabungkan pada iterasi pertama akan digabungkan kembali pada iterasi ke-3. Pada iterasi ke-3, cluster yang telah digabungkan pada iterasi ke-1 dan cluster 4 akan digabungkan kembali. Cluster yang akan dihasilkan berikutnya akan muncul pada iterasi ke-3. Begitu seterusnya hingga iterasi ke-19.

Jumlah cluster yang terbaik dapat dilihat dari selisih koofisien dari setiap iterasinya.

Jumlah Cluster	Tahap Terakhir Agglomeration	Koofisien Tahap Ini	Selisih
2	12.00	11.75	0.25
3	11.75	9.95	1.80
4	9.95	9.49	0.46
5	9.49	9.43	0.06

Gambar 5. Selisih Setiap Iterasi

Selisih terbesar yaitu 1.80 sehingga jumlah *cluster* terbaik adalah 3 *cluster*. Maka, dendrogram hasil pengelompokan berdasarkan ukuran akan dipotong seperti berikut :



Gambar 6. Pemotongan Dendrogram

Hasil akhir dari pengelompokan ini terbentuk 3 *cluster*, masing-masing kelompok sebagai berikut :

Tabel 1. Anggota Cluster 1

ID Distributor	Distributor
6	Renny
7	Hani
11	Jay Ezah
12	Heppi
20	Syanah

Tabel 2. Anggota Cluster 2

ID Distributor	Distributor
2	Kurniawan
4	Eni
14	Cut Opie
19	Mike Sariang

Tabel 3. Anggota Cluster 3

ID Distributor	Distributor
1	Rani
3	Wulandari
5	Poppy
8	Endri
9	Ibeth
10	Eerstiana
13	Zulaikha
15	Ema
16	Lina
17	Misrah
18	Dewi

Dari hasil pengelompokan distributor berdasarkan penjualan sepatu didapatkan hasil penjualan tertinggi dari setiap *cluster* yaitu sebagai berikut :

Tabel 4. Penjualan Tertinggi Pada Setiap Cluster

Cluster	Motif dan Ukuran	Jumlah
Cluster 1	QZ_40	28
Cluster 2	HB_38	16
	HB_39	16
	PD_39	16
Cluster 3	RV_42	10
	HB_40	10

Pada cluster 1 penjualan tertinggi yang sepatu dengan motif QZ dengan nomor 40. Pada cluster 2 penjualan tertinggi yaitu motif HB dengan nomor sepatu 38 dan 39 serta motif PD dengan nomor sepatu 39. Sedangkan pada cluster 3 penjualan tertinggi yaitu motif RV dengan nomor 42 dan motif HB dengan nomor 40.

3. PENUTUP

Hasil dari penelitian yang dilakukan diperoleh kesimpulan bahwa aplikasi pembentukan kelompok distributor yang telah dibangun, dapat membantu manajer pemasaran untuk membentuk kelompok distributor yang sesuai dengan kemiripan penjualan produk sehingga dapat dijadikan acuan untuk pendistribusian dan strategi bisnis yang tepat.

Ada beberapa saran yang dapat dilakukan untuk pengembangan aplikasi pembentukan kelompok distributor ini, antara lain :

1. Pembuatan motif baru dilakukan CV. XYZ dalam jangka waktu yang cukup lama. Maka dari itu, untuk mengantisipasinya perlu penambahan fitur pengaturan atribut, sehingga jika terdapat motif baru pembentukan kelompok dapat disesuaikan dengan penambahan motif baru.

2. Metode pengelompokan dapat dikembangkan lagi dengan menggunakan algoritma lain atau menggabungkan 2 algoritma yang berbeda, seperti menggabungkan algoritma AHC (*agglomerative hierarchical clustering*) dengan algoritma K-Means, sehingga dapat dibandingkan mana algoritma yang lebih baik dalam melakukan pengelompokan.

DAFTAR PUSTAKA

- [1] S. B, Data Mining Teknik Pengumpulan Data Untuk Keperluan Bisnis, Yogyakarta: Graha Ilmu, 2007.
- [2] "CRISP-DM 1.0 Step-by-step data mining guide," 2000.
- [3] Fathansyah, Basis Data, Informatika, 2015.
- [4] R. Ramakrishnan dan J. Gehrke, Database Management System Third Edition, Singapore, 2003.
- [5] J. Han dan M. Kamber, Data Mining : Concepts and Techniques, 2006.
- [6] T. Badriyah, "Penanganan Missing Value dan Outlier".
- [7] Sunitha, M.BalRaju, J.Sasikiran dan V. R. , "Automatic Outlier Identification in Data Mining Using IQR in Real-Time Data," International Journal of Advanced Research in Computer and Communication Engineering, vol. 3, no. 6, 2014.
- [8] B. Susanto, "Data Preprocessing," Versi 1.2, 2013.
- [9] P. Berkhin, "Survey of Clustering Data Mining Techniques".
- [10] F. A. Hermawati, Data Mining, Yogyakarta: Andi, 2013.
- [11] R. Xu dan D. Wunsch, Clustering.
- [12] M. Caccam dan J. Refran, "Cluster Analysis".
- [13] A. Nugroho, Rekayasa Perangkat Lunak Berorientasi Objek dengan Metode USDP, Yogyakarta: Andi, 2010.
- [14] Suryana, Model Praktis Penelitian Kuantitatif dan Kualitatif, 2010.