PENERAPAN *DATA MINING* PADA PENJUALAN ROTI KEMASAN MENGGUNAKAN METODE *CLUSTERING* DI ROTI RIKI

Muhamad Sandy Hasanudin[1], Dian Dharmayanti[2]

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

ABSTRAK

Roti Riki merupakan sebuah unit usaha yang bergerak di bidang produksi olahan tepung. Pada kegiatan operasionalnya, Roti Riki memproduksi produk olahan tepung yaitu roti manis isi dengan berbagai varian isi. Roti Riki tidak mendistribusikan langsung terhadap penjual eceran atau warung-warung. Distributor-distributorlah yang akan memasok ke warung-warung yang melayani pelanggan secara langsung. Distributor mengambil produk langsung ke gudang. Distributor tidak dapat menentukan varian produk yang dapat diambil. Mereka hanya mengambil sesuai stok yang ada di gudang. Pada proses pendistribusian seperti itu diketahui kurang maksimal. Hal itu terjadi dikarenakan pendistribusian varian roti tidak sesuai dengan kebutuhan distributor. Akibatnya adalah pengembalian sejumlah produk oleh distributor.
Berdasarkan pemaparan masalah sebelumnya, solusi yang ditawarkan untuk mengatasi permasalah tersebut adalah membentuk pengelompokan distributor berdasarkan data transaksi. Hasil dari pengelompokan tersebut dapat menjadi pengetahuan dengan menganalisis karakteristik yang berbeda yang memungkinkan dibutuhkan varian atau pemasaran yang berbeda dari setiap kelompoknya. Pembentukan kelompok distributor dapat dilakukan dengan menggunakan metode clustering di data mining. Salah satu algoritma untuk melakukan clustering (pengelompokan) adalah algoritma Agglomerative Hierarchical Clustering (AHC). Algoritma AHC mempunyai kemampuan untuk membentuk cluster-cluster berdasarkan kedekatan jarak antar data.
Berdasarkan pengujian telah dilakukan dapat disimpulkan bahwa aplikasi yang dibangun ini membantu pihak pemasaran Roti Riki melakukan pemasaran dan strategi bisnis yang tepat untuk setiap kelompok distributornya sesuai dengan hasil analisis.
Kata kunci : *distributor, data mining, clustering,*

1. PENDAHULUAN

Roti Riki merupakan sebuah unit usaha yang bergerak di bidang produksi olahan tepung. Pada kegiatan operasionalnya, Roti Riki memproduksi produk olahan tepung yaitu roti manis isi dengan berbagai varian isi seperti meses, mocca, coklat, kopyor, blueberry dan Roti Riki pun memproduksi kue pia dengan isi kacang ijo, pia coklat. Jenis roti yang diproduksi yaitu roti manis kemasan yang sering didapati di warung-warung di pinggir jalan. Roti Riki tidak mendistribusikan langsung terhadap penjual eceran atau warung-warung. Distributor-distributorlah yang akan memasok ke warung-warung yang melayani pelanggan secara langsung. Setiap produk yang diambil oleh para distributor, Roti Riki biasanya mencatat data transaksi beserta jumlah produk yang diambil ke dalam database.
Setelah melakukan wawancara dengan ketua bagian produksi, didapatkan informasi bahwa pihak Roti Riki terkadang menambahkan varian isi rasa pada produk-produknya. Biasanya, Roti Riki membuat inovasi untuk memperbanyak pilihan bagi para distributor untuk memperluas pangsa pasar. Inovasi yang dilakukan seperti menambah varian isi dengan rasa baru dan juga memperbesar ukuran roti yang tentunya ditawarkan dengan harga yang variatif. Sayangnya tidak selalu varian rasa isi baru dengan ukuran tertentu yang diproduksi itu laku dipasaran. Apabila roti-roti yang telah didistribusikan dalam jangka waktu tertentu tidak laku bahkan basi, maka distributor-distributor biasanya mengembalikan produk-produk yang tidak laku tersebut ke pabrik Riki Roti. Pihak Pemasaran Riki Roti ingin mengurangi potensi pengembalian seminimal mungkin dan menawarkan varian-varian produk sesuai dengan kebutuhan distributor. Sebab tiap distributor mempunya kebutuhan terhadap produk yang berbeda.
Berdasarkan hasil pengamatan terhadap masalah yang telah dipaparkan sebelumnya, solusi yang mungkin dilakukan adalah membentuk kelompok distributor sehingga karakteristik kebutuhan produk dari setiap distributor dapat diketahui [1]. Kebutuhan setiap distributor tersebut dapat dilihat berdasarkan data transaksi yang telah dilakukan. Maka dari itu segmentasi akan dilakukan berdasar-
kan data transaksi.
Hasil pengamatan dari data transaksi ternyata tiap data memiliki pola data dan memiliki keterkaitan yang menunjukkan bahwa ada kejadian berulang. Untuk mendapatkan informasi tersebut secara lebih cepat dan efisien diperlukan suatu bantuan teknologi

informasi, dalam hal ini yaitu Data Mining [2]. Data Mining adalah suatu proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan, dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [3]. Dalam data mining terdapat beberapa metode yaitu prediksi, klasifikasi, pengklusteran, dan asosiasi. Salah satu metode tersebut adalah metode *clustering*, yaitu metode pengelompokan sejumlah data atau objek ke dalam *cluster* (kelompok) sehingga setiap dalam cluster tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam *cluster* yang lainnya [4].

Berdasarkan permasalahan yang telah dipaparkan sebelumnya, pada penelitian ini akan dibangun aplikasi *Data Mining* yang menerapkan metode *Clustering* yang berguna untuk mensegmentasi distributor-distributor di Roti Riki yang berguna sebagai informasi untuk rekomendasi penawaran produk-produk kepada distributor.

2. ISI PENELITIAN

Pada bagian ini akan dibahas mengenai tahapan penelitian yang dilakukan.

2.1 Pemahaman Bisnis

Tujuan bisnis dari Roti Riki adalah menjadi produsen roti kemasan yang dapat melayani distributor-distributor sesuai dengan kebutuhan mereka. Adapun menjadi faktor penting agar tujuan bisnis Roti Riki ini tercapai yaitu dengan menyediakan dan mendistribusikan produk yang sesuai dengan permintaan dan karakteristik distributor.

Sasaran dari penerapan data mining pada penjualan produk di Roti Riki ini adalah untuk mengelompokkan distributor-distributor berdasarkan data transaksi.

Penilaian situasi adalah penulisan fakta-fakta yang ada di tempat penelitian secara rinci. Fakta tersebut meliputi ketersediaan sumber daya, kebutuhan yang diperlukan, dan keuntungan yang akan didapat dari penelitian *data mining*. Berikut merupakan hasil dari kegiatan penilaian situasi di Roti Riki.

a) Ketersediaan Sumber Daya

Pada kegiatan ini mencatat berbagai sumber daya yang tersedia untuk kebutuhan penelitian. Sumber daya tersebut meliputi perangkat keras, perangkat lunak, data, dan personel. Sumber daya yang tersedia di Roti Riki untuk penelitian ini adalah sebagai berikut :

1. Ketersediaan Perangkat Keras dan Perangkat Lunak

Berikut adalah tabel yang menerangkan ketersediaan perangkat keras di Roti Riki yang akan digunakan untuk menjalankan perangkat lunak yang dibutuhkan untuk penelitian.

Tabel 1 Perangkat Keras Yang Tersedia

No	Perangkat Keras	Spesifikasi
1	*Processor*	Intel Core i3 2102 3.0 GHz
2	*VGA*	Intel HD 3000 1024 MB
3	*Memory*	RAM 2GB
4	*Harddisk*	500 GB
5	*Monitor*	LCD 14 *inch*
6	*Keyboard*	*Standard Port* PS2
7	*Mouse*	*Standard Port* USB

Sedangkan Tabel 2 merupakan perangkat lunak yang tersedia yang akan digunakan untuk menjalankan aplikasi pengolahan data mining, membangun sistem hasil analisis, dan membuat laporan penelitian.

Tabel 2 Perangkat Lunak yang tersedia

No	Perangkat Lunak	Spesifikasi
1	Sistem Operasi	Microsoft Windows 7
2	Program	a. Microsoft Office 2013 b. Net Framework 3.5 c. Xampp d. R Tools

2. Ketersediaan Data

Data yang tersedia untuk penelitian ini adalah data transaksi. Data tersebut dikumpulkan dan diolah oleh kepala bagian pemasaran. Untuk keperluan penelitian, peneliti telah mendapat izin dari kepala bagian pemasaran untuk mengakses dan mengolah data tersebut.

3. Ketersediaan Personel

Personel yang tersedia berasal dari internal dan eksternal Roti Riki. Pada penelitian ini, Kepala Bagian Pemasaran Roti Riki adalah satu-satunya personel yang terkait. Kepala Bagian Pemasaran Roti Riki yang menyediakan data untuk dianalisis pada penelitian *data mining* ini.

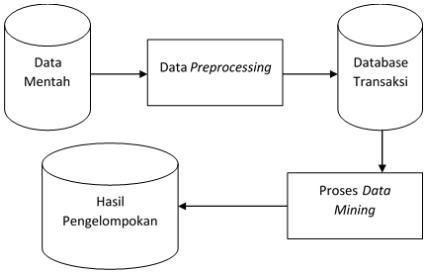
b) Kebutuhan yang diperlukan

Kebutuhan yang diperlukan pada penelitian kali ini adalah data distributor dan data transaksi selama bulan Mei 2016 di Roti Riki.

c) Keuntungan yang didapat dari penelitian *data mining*.

Keuntungan yang didapat dari penelitian data mining ini adalah adanya informasi dan pengetahuan baru bagi pihak pemasaran Roti Riki dalam hal ini karakteristik kelompok dari distributor berdasarkan data transaksi. Sehingga informasi tersebut bisa digunakan sebagai salah satu rekomendasi untuk membuat strategi bisnis yang tepat terhadap distributor di Roti Riki.

Perencanaan proyek adalah tahapan-tahapan yang dilalui untuk mencapai tujuan dari sasaran *data mining*. Berikut tahapan perencanaan proyek untuk mencapai sasaran *data mining* :



Gambar 1 Perencanaan Proyek

Data mentah yang didapatkan dari data transaksi selama bulan Mei 2016 di Roti Riki. Data mentah tersebut menjadi masukan bagi sistem. Pada sistem dilakukan *Preprocessing* terhadap data tersebut, kemudian menjadi sebuah dataset baru dan dimasukkan kedalam sebuah penyimpanan *database*. Data yang berada di dalam *database* tersebut barulah dilakukan proses *mining* yang kemudian menghasilkan hasil pengelompokan.

2.2 Pemahaman Data

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

2.2.1 Pengumpulan Data Awal

Data dalam penelitian adalah data transaksi di Roti Riki yang berformat file *Microsoft excel* (.xlsx) yang nantinya akan dilakukan proses *mining* pada sebuah aplikasi. Berikut ini merupakan sebagian data awal yang berupa data transaksi pada bulan Mei 2016 .

ID	Nama	Alamat	Roti Manis 1000				Kue Pia		Roti Manis 2000				Sisa Roti Manis 1000				Sisa Kue Pia		Sisa Roti Manis 2000			
			Meses	Mocca	Coklat	Blueberry	Kopyor	Kacang Ijo	Coklat	Meses	Kombinasi	Coklat	Kopyor	Meses	Mocca	Coklat	Blueberry	Kopyor	Kacang Ijo	Coklat	Meses	Kombinasi	Coklat	Kopyor
RRD-01	Toni Amajaya	Purwakarta	7490	5830	4590	6502	4215	6400	2640	3320	3460	4170	3230	0	0	0	0	0	0	0	0	0	0	0
RRD-02	Asep Rachman	Rance Ekek	3210	2750	4550	300	950	1330	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RRD-03	Aris Winarso	Baie Endah	2230	2230	3280	1050	1010	2280	860	585	820	885	410	0	0	0	5	0	5	0	0	0	0	0
RRD-04	Goyon Winarno	Cicalengka	3100	2535	3900	570	810	2330	550	440	290	685	190	289	320	422	15	35	267	65	0	0	37	0
RRD-05	Abas Yahya	Majalaya	6970	6040	4380	5590	4120	8470	2990	2555	2340	2900	2220	0	0	0	0	0	0	0	0	0	0	0
RRD-06	Mamo Surono	Majalaya	4670	3680	3500	1420	1800	3410	840	640	455	630	180	20	50	15	0	55	30	0	0	0	15	15
RRD-07	Aceng Sanjaya	Cibiru	3890	4320	3580	1700	1390	3420	870	825	200	885	290	24	124	0	0	0	32	0	0	0	0	0
RRD-08	Tursino Widagdo	Majalaya	3710	3500	5130	1800	1120	3870	1420	1215	800	1430	720	7	5	0	13	1	0	0	0	10	0	14
RRD-09	Imas Darmaji	Majalaya	3615	3340	3920	2740	2285	2925	2075	2115	1925	2255	2035	0	0	0	0	0	0	0	0	0	0	0
RRD-10	Ferdi Septian	Garut	3630	3080	4470	1675	1310	3255	1170	870	930	1225	550	320	190	199	66	44	225	24	15	13	11	0
RRD-11	Sapriol Mahmud	Majalaya	4625	4585	5910	1445	1480	4015	810	720	475	860	365	0	47	0	0	0	28	0	35	0	0	0
RRD-12	Toyo Faturakhmat	Tasik	10150	11650	6290	7730	5270	7820	3585	4540	4755	5225	4530	10	5	0	5	0	22	0	0	0	0	8
RRD-13	Gilang Rambu	Garut	3980	3295	5250	1400	1770	3515	1440	830	920	1045	635	89	0	82	20	0	35	33	0	26	15	0
RRD-14	Suyoto	Sumedang	4960	3745	6345	1985	1710	4525	1870	1385	1210	1695	850	123	330	162	40	0	345	220	0	20	0	0
RRD-15	Atam Gunardi	Karawang	7530	8780	5740	9660	5850	9040	5160	4540	4070	5030	4100	0	0	0	0	0	0	0	0	0	0	0
RRD-16	Wasko Satrio	Lembang	10170	11600	4355	7805	4770	7770	2992	3150	2495	3650	2420	221	356	230	155	0	431	0	0	189	224	52
RRD-17	Yono Fathah	Sumedang	4880	3850	6090	2430	2285	5210	1795	1255	1190	2040	925	122	0	114	0	0	123	0	15	0	0	76
RRD-18	Lusiyno Hadi Kus	Rance Ekek	3160	2025	3850	1175	1330	2340	640	425	210	580	305	65	22	221	267	188	0	6	0	3	7	21
RRD-19	Chong Sirait	Sumedang	3870	3200	4640	1550	1790	3905	1420	1095	965	1475	640	0	0	0	0	0	0	0	0	0	0	0
RRD-20	Tino Wijaya	Cilacap	2900	2300	3110	1450	1250	2270	950	630	530	590	430	0	0	0	0	0	0	0	0	0	0	0

Gambar 2. Data Transaksi Bulan Mei2016

2.2.2 Penjelasan Data

Data awal yang telah dikumpulkan akan dijelaskan pada tahap ini. Pada data terdapat 25 atribut, yaitu ID, Nama, Alamat, Roti Manis 1000 Meses, Roti Manis 1000 Mocca, Roti Manis 1000 Coklat, Roti Manis 1000 Kopyor, Roti Manis 1000 Blueberry, Kue Pia Coklat, Kue Pia Kacang Ijo, Roti Manis 2000 Meses, Roti Manis 2000 Kombinasi, Roti Manis 2000 Coklat, Roti Manis 2000 Kopyor, Sisa Roti Manis 1000 Meses, Sisa Roti Manis 1000 Mocca, Sisa Roti Manis 1000 Coklat, Sisa Roti Manis 1000 Kopyor, Sisa Roti Manis 1000 Blueberry, Sisa Kue Pia Coklat, Sisa Kue Pia Kacang Ijo, Sisa Roti Manis 2000 Meses, Sisa Roti Manis 2000 Kombinasi, Sisa Roti Manis 2000 Coklat, Sisa Roti Manis 2000 Kopyor

Tabel 3 Penjelasan Atribut Data Transaksi

Keterangan		
Fungsi	Untuk mengetahui produk apa saja yang dibeli oleh distributor serta jumlah produk yang dikembalikan	
Atribut	ID	ID Distributor
	Nama	Nama dari pemilik distributor
	Alamat	Alamat dari distributor
	Roti Manis 1000 Meses, Roti Manis 1000 Mocca, Roti Manis 1000 Coklat, Roti Manis 1000 Kopyor, Roti Manis 1000 Blueberry, Kue Pia Coklat, Kue Pia Kacang Ijo, Roti	Jenis Produk atau Sisa Produk

Keterangan		
	Manis 2000 Meses, Roti Manis 2000 Kombinasi, Roti Manis 2000 Coklat, Roti Manis 2000 Kopyor, Sisa Roti Manis 1000 Meses, Sisa Roti Manis 1000 Mocca, Sisa Roti Manis 1000 Coklat, Sisa Roti Manis 1000 Kopyor, Sisa Roti Manis 1000 Blueberry, Sisa Kue Pia Coklat, Sisa Kue Pia Kacang Ijo, Sisa Roti Manis 2000 Meses, Sisa Roti Manis 2000 Kombinasi, Sisa Roti Manis 2000 Coklat, Sisa Roti Manis 2000 Kopyor	

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

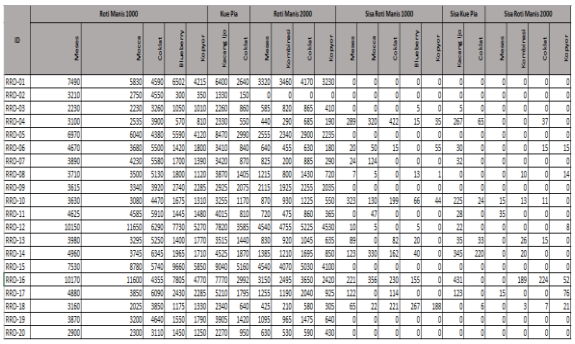
Pemilihan data dilakukan untuk menentukan data yang akan digunakan pada tahap pemodelan atau proses *clustering*. Pemilihan data tersebut meliputi pemilihan atribut dan *record* yang akan ada dalam sistem. terdapat 23 atribut yang akan dipilih untuk tahap pemodelan yaitu ID, Roti Manis 1000 Meses, Roti Manis 1000 Mocca, Roti Manis 1000 Coklat, Roti Manis 1000 Kopyor, Roti Manis 1000 Blueberry, Kue Pia Coklat, Kue Pia Kacang Ijo, Roti Manis 2000 Meses, Roti Manis 2000 Kombinasi, Roti Manis 2000 Coklat, Roti Manis 2000 Kopyor, Sisa Roti Manis 1000 Meses, Sisa Roti Manis 1000 Mocca, Sisa Roti Manis 1000 Coklat, Sisa Roti Manis 1000 Kopyor, Sisa Roti Manis 1000 Blueberry, Sisa Kue Pia Coklat, Sisa Kue Pia Kacang Ijo, Sisa Roti Manis 2000 Meses, Sisa Roti Manis 2000 Kombinasi, Sisa Roti Manis 2000 Coklat, Sisa Roti Manis 2000 Kopyor.

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 pengembangan set data baru atau memproduksi atribut turunan melalui transformasi data yang sesuai dengan kebutuhan pemodelan. Menurunkan atribut-atribut yang diperlukan pada tahap pemodelan dan membuang atribut yang tidak diperlukan untuk tahap pemodelan. Berikut hasil penyusunan data-data tersebut :



Gambar 3. Penyusunan Data

2.3.4 Integrasi Data

Pada tahap ini adalah data yang akan digunakan dalam tahapan pemodelan adalah data transaksi pada bulan Mei 2016 yang telah mengalami proses perubahan atribut, berikut beberapa atribut yang diganti :

a. Roti Manis 1000 menjadi Roti1
b. Roti Manis 2000 menjadi Roti2
c. Kue Pia menjadi Pia
d. Sisa Roti Manis 1000 menjadi Sroti1
e. Sisa Roti Manis 2000 menjadi Sroti2
f. Sisa Kue Pia menjadi Spia

Kemudian hasil perubahan tersebut digabungkan dengan atribut varian isi dihubungkan oleh tanda *underscore* diantara dua atribut tersebut. Setelahnya dilakukan penggabungan dua atau lebih tabel yang memiliki informasi berbeda tentang objek yang sama ke dalam set data baru..

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 hierarchical clustering, dan model yang akan digunakan adalah algoritma *Agglomerative Hierarchical Clustering* (AHC).

2.4.2 Mekanisme Pemodelan

Pada tahap ini mejelaskan 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 kelompok. Jika n = jumlah data dan k = jumlah kelompok, maka k=n.
2. Menghitung matriks jarak antar kelompok (menggunakan *euclidean distance*).
3. Cari dua kelompok yang mempunyai jarak paling dekat (*complete linkage*) lalu gabungkan kedua kelompok tersebut sehingga k =k -1.
4. Jika k >1, ulangi langkah 2 dan 3.
5. Jika k = 1, perulangan selesai.
6. Menghasilkan kelompok dengan memotong dendrogram pada tingkat yang tepat.

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

Ubah setiap objek data menjadi kelompok

Inisialisasi kelompok awal adalah sebanyak 20 kelompok, dikarenakan data (n) = kelompok (k), kemudian dihitung jarak antara kelompok dengan kelompok yang ada dengan menggunakan persamaan Euclidean Distance.

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

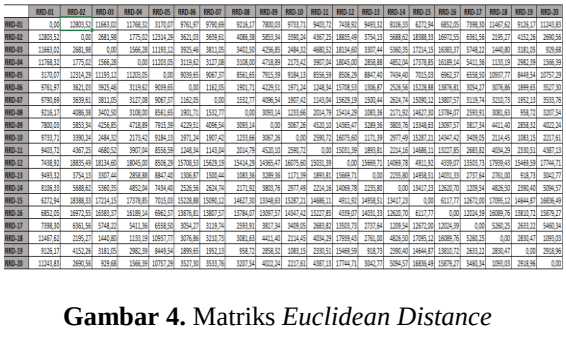
Setiap *cluster* akan dihitung jarak dirinya sendiri dengan *cluster* lainnya. Berikut ini merupakan proses perhitungan jarak antar *cluster*. Berikut ini merupakan contoh perhitungan jarak berdasarkan ukuran :

$$d(RRD-01, RRD-20) = \sqrt{\begin{array}{c}(7490-2900)^2+(5830-2300)^2+(4590-3110)^2\\(6502-1450)^2+(4215-1250)^2+(6400-2270)^2+\\(2640-950)^2+(3320-630)^2+(3460-530)^2\\+(4170-590)^2+(3230-430)^2+(0-0)^2+\\(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2\\+(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2\\+(0-0)^2\end{array}} = 11243,83$$

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

Dalam tahap ini, setiap kelompok akan dihitung jarak dengan dirinya sendiri dan jarak dengan kelompok lainnya. Berikut ini merupakan proses perhitungan jarak antar kelompok dengan menggunakan persamaan *Euclidean Distance*. Berikut ini adalah contoh perhitungan jarak berdasarkan jumlah transaksi :

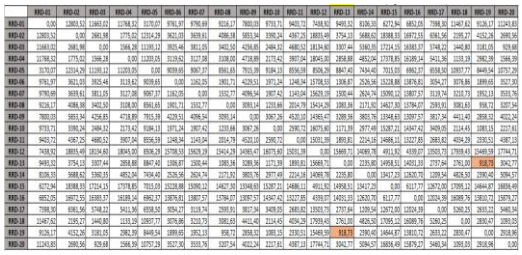


Gambar 4. Matriks *Euclidean Distance*

Langkah 3

Mencari dua kelompok yang saling berdekatan.

Pada tahapan ini akan dihitung jarak antara dua kelompok dengan menggunakan complete linkage. Dengan menggunakan metode ini, maka akan dicari jarak yang paling dekat dari dua kelompok. Berikut adalah contoh dua kelompok yang saling berdekatan.



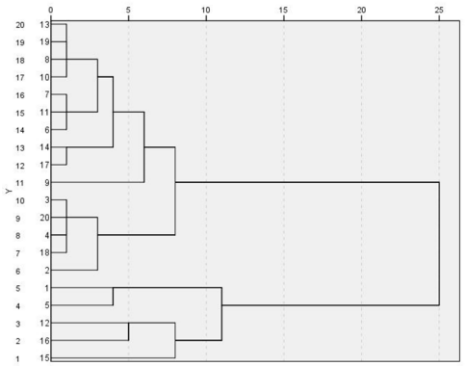
Gambar 5. Matriks *Euclidean Distance* Iterasi ke-1

Langkah ke-4

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-5

Ketika jumlah *cluster* berjumlah 1,maka tahap 3 telah selesai dan hasil akhir dari perulangan berdasarkan transaksi berikut ini merupakan *dendogram* dari hasil pengelompokkan.seperti terlihat pada Gambar 6. Hasil Akhir *Dendogram*.



Gambar 6. 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 7.

Iterasi	Kombinasi Kelompok		Koefisien	Iterasi Kombinasi Kelompok sebelumnya		Iterasi Selanjutnya
	Kelompok 1	Kelompok 2		Kelompok 1	Kelompok 2	
1	RRD-13	RRD-19	918,727	0	0	3
2	RRD-03	RRD-20	929,677	0	0	9
3	RRD-08	RRD-13	1083,150	0	1	7
4	RRD-4	RRD-18	1133,188	3	0	9
5	RRD-07	RRD-11	1143,042	2	0	8
6	RRD-14	RRD-17	1209,541	5	0	13
7	RRD-08	RRD-10	1233,660	0	0	10
8	RRD-06	RRD-07	1248,344	0	7	10
9	RRD-03	RRD-04	1566,393	0	0	11
10	RRD-06	RRD-08	2590,716	8	4	13
11	RRD-02	RRD-03	2690,558	0	6	17
12	RRD-01	RRD-05	3170,066	11	10	18
13	RRD-06	RRD-14	3409,053	12	9	15
14	RRD-12	RRD-16	4339,074	0	0	16
15	RRD-06	RRD-09	4520,096	13	0	17
16	RRD-12	RRD-15	6117,774	0	0	18
17	RRD-02	RRD-06	6361,557	16	0	19
18	RRD-01	RRD-12	8506, 293	14	15	19
19	RRD-01	RRD-02	18835,490	18	17	0

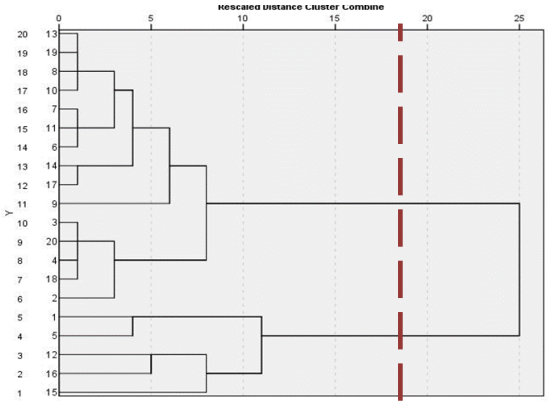
Gambar 7. *Agglomeration Schedule*

Jumlah *cluster* yang terbaik dapat dilihat dari selisih koefisien dari setiap iterasinya. Selisih dari koefisien segmentasi berdasarkan jumlah transaksi dapat dilihat pada Gambar 8.

Jumlah Kelompok	Terakhir *Agglome*	Koefisien Tahap ini	Selisih
2	18835,490	8506,293	10329,197
3	8506,293	6361,557	2144,736
4	6361,557	6117,774	243,783
5	6117,774	4520,096	1597,678
6	4520,096	4339,074	181,022
7	4339,074	3409,053	930,021
8	3409,053	3170,066	238,987
9	3170,066	2690,558	479,508
10	2690,558	2590,716	99,842
11	2590,716	1566,393	1024,323
12	1566,393	1248,344	318,049
13	1248,344	1233,660	14,684
14	1233,660	1209,541	24,119
15	1209,541	1143,042	66,499
16	1143,042	1133,188	9,854
17	1133,188	1083,150	50,038
18	1083,150	929,677	153,473
19	929,677	918,727	10,950

Gambar 8. Selisih Koefisien Pengelompokan

Selisih terbesar yaitu 4,37 sehingga jumlah *cluster* terbaik adalah 2 *cluster*. Maka *dendogram* hasil pengelompokkan berdasarkan jumlah transaksi akan dilakukan pemotongan seperti terlihat pada Gambar 8.



Gambar 9. Pemotongan *Dendogram*

Hasil akhir dari pengelompokan ini, terbentuk 2 kelompok dengan masing-masing kelompok dapat dilihat pada Tabel 3 dan Tabel 4.

Tabel 2. *Cluster* 1

ID	Nama Pemilik Distributor	Alamat
RRD-02	Asep Rachman	Ranca Ekek
RRD-03	Aris Winarso	Bale Endah
RRD-04	Goyon Winarno	Cicalengka
RRD-06	Marno Suroto	Majalaya
RRD-07	Ateng Sanjaya	Cibiru
RRD-08	Tursino Widagdo	Majalaya
RRD-09	Imas Darmaji	Majalaya
RRD-10	Ferdi Septian	Garut
RRD-11	Saprol Mahmud	Majalaya
RRD-13	Gilang Rambu	Garut
RRD-14	Suyoto	Sumedang
RRD-17	Yono Fathah	Sumedang
RRD-18	Lusyuno Hadi Kusumo	Ranca Ekek
RRD-19	Otong Sirait	Sumedang
RRD-20	Tino Wijaya	Cilacap

Tabel 3. *Cluster* 2

ID	Nama Pemilik Distributor	Alamat
RRD-01	Toni Atmajaya	Purwakarta
RRD-05	Abas Yahya	Majalaya
RRD-12	Toyo Faturakhman	Tasik
RRD-15	Atam Gunardi	Karawang
RRD-16	Wasko Satrio	Lembang

Dari hasil pengelompokan distributor, didapatkan pengembalian terendah

Tabel 4. Pengembalian Terendah

Kelompok	Jenis Produk
Kelompok 1	Sroti2_Meses
	Sroti2_Kombinasi
	Sroti2_Coklat
Kelompok 2	Spia_Coklat
	Sroti1_Kopyor
	Sroti2_Meses

Dari hasil pengembalian terendah tersebut, dapat dihasilkan rekomendasi produk untuk setiap *cluster* yang dapat dilihat pada Tabel 5 berikut :

Tabel 5. Rekomendasi Produk Setiap Kelompok

Kelompok	Rekomendasi
Kelompok 1	Mengutamakan penjualan Roti 2000 Meses, Roti 2000 Kombinasi, Roti 2000 Coklat.
Kelompok 2	Mengutamakan penjualan Kue Pia Coklat, Roti 1000 Kopyor, Roti 2000 Meses.

3. PENUTUP

Hasil dari penelitian yang telah dilakukan, diperoleh kesimpulan bahwa aplikasi pengelompokan distributor yang telah dibangun dapat membantu manager pemasaran untuk membentuk kelompok yang sesuai dengan kemiripan transaksi produk sehingga dapat dijadikan acuan untuk pendistribusian dan strategi bisnis yang tepat untuk karakteristik setiap distributornya.

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

1. Pembuatan varian baru dilakukan Roti Riki dalam jangka waktu yang cukup lama. Maka dari itu, untuk mengantisipasinya perlu penambahan fitur pengaturan atribut, sehingga jika terdapat varian baru, pembentukan kelompok dapat disesuai dengan penambahan varian baru.
2. Metode pengelompokkan dapat dikembangkan lagi dengan menggunakan algoritma lain atau menggabungkan 2 algoritma yang berbeda, seperti menggabungkan algoritma AHC (A*gglomerative Hierarchical Clustering*) dengan algoritma K-Means, sehingga dapat dibandingkan mana algoritma yang lebih baik dalam pengelompokkan.
3. Metode AHC yang digunakan dapat dikembangkan lagi dengan menggunakan Average linkage, sehingga dapat dibandingan mengan metode lain mana yang lebih akurat untuk menghitung jarak.
4. Metode yang digunakan dalam perhitungan jarak bisa dikembangkan lagi dengan menggunakan metode Manhattan, sehingga dapat dibandingkan dengan metode perhitungan yang lain.

DAFTAR PUSTAKA

[1] D. S. Rajagopal, "Customer Data Clustering Using Data Mining Technique," International Journal of Database Management Systems (IJDMS), vol. 3, no. 4, pp. 1-3, 2011.

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

[3] D. T. Larose, Discovering knowledge in data, New Jersey: John Wiley & Sons, 2005.

[4] F. A. Hermawati, Data Mining, Yogyakarta: Penerbit Andi, 2013.
[5] R. W. P. Chapman, 2000. [Online]. Available: http://the- modeling-agency.com/crisp-dm.pdf. [Diakses 23 Agustus 2016].
[6] Fathansyah, Basis Data, Bandung: Informatika, 2015.
[7] K. Tampubolon, H. Saragih dan B. Reza, "Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-alat Kesehatan," Majalah Ilmiah Informasi dan Teknologi Ilmiah (INTI), vol. 1, no. 1, 2013.
[8] L. (Allium ursinum L., "Research Gate," November 2015. [Online]. Available: https://www.researchgate.net/figure/284724792_fig2_Fig-2-A-Dendrogram-AHC-analysis-representing-dissimilarity-relationships-of-14. [Diakses 29 October 2016].
[9] I. Sommerville, Software Engineering (9th Edition), Pearson, 2010.
[10] I. Supardi, Semua Bisa Menjadi Programming Java Basic Programming, Jakarta: Elex Media, 2010.
[11] B. Hambling dan P. v. Goathem, User Acceptance Testing : Step by Step Guide, Swindon: BCS Institute, 2013.
[12] T. Alfina, B. Santosa dan A. R. Barakbah, "Analisa Perbandingan Metode Hierarchical," JURNAL TEKNIK ITS, vol. 1, p. 1, 2012.
[13] R. Ramakrishnan dan J. Gehrke, Database Management System, Singapore, 2003.