Penerapan Data Mining Pada Data Penjualan Produk Untuk Membentuk Segmentasi *Outlet* di Blackjack Industries Bandung

Faza Rahthin A[1], Dian Dharmayanti [2]
[1,2]Informatika - Universitas Komputer Indonesia
Jl. Dipatiukur 112-114 Bandung
E-mail : fazarahthin@gmail.com[1], diandmdr@yahoo.com[2]

ABSTRAK

Blackjack Industries merupakan sebuah produsen pakaian jadi yang meliputi t-shirt, kemeja dan tas yang memasarkan produknya pada *outlet-outlet* yang tersebar di beberapa wilayah Indonesia. Blackjack Industries mengirimkan produk baru setiap sebulan sekali ke setiap *outlet* dan *outlet* menerima produk tersebut lalu memasarkan kepada pelanggan. Produk baru yang dimaksud adalah kategori produk jenis dan model. Berdasarkan survey kuesioner yang telah dilakukan terhadap 20 *outlet* Blackjack Industries diketahui bahwa 14 *outlet* mengalami kendala dalam hal ketidaksesuaian kategori produk jenis dan model yang artinya bahwa kategori produk jenis dan model yang dijual pada setiap *outlet* tidak sesuai dengan kebutuhan pelanggan. Hal ini menyebabkan banyak kategori produk jenis dan model yang tidak laku terjual. Maka kategori produk jenis dan model yang dibutuhkan tiap *outlet* berbeda dan membutuhkan karakteristik yang tepat untuk setiap *outlet*.

Ketersediaan data, kebutuhan akan sebuah informasi (atau pengetahuan) yang berguna sebagai pendukung pengambilan keputusan. Data yang didapat berguna dalam mendapatkan karakteristik yang tepat dari setiap *outlet*. Karakteristik di dapat dengan melakukan analisis terhadap data penjualan produk terhadap masing-masing *outlet* dengan menggunakan Data Mining. Salah satu metode yang terdapat dalam data mining yang digunakan dalam penelitian ini adalah *cluster*ing. dimana teknik yang dipakai menggunakan algoritma Agglomerative hierarchical *cluster*ing (AHC) dalam pembentukan segmentasi sesuai karakteristik yang di dapat dari data penjualan.

Berdasarkan hasil pengujian yang telah dilakukan dapat disimpulkan bahwa segmentasi *outlet* ini dapat membantu pihak Blackjak Industries dalam menentukan kebutuhan tiap *outlet* berdasarkan karakteristik yang di dapat dan Mempermudah *manager marketting* dalam menentukan rekomendasi produk yang tepat.

Kata kunci: *Hierarchical Agglomerative Clustering, Data Mining, Clustering*, Segmentasi, *Outlet*

1. PENDAHULUAN

1.1 Latar Belakang

Blackjack Industries merupakan sebuah produsen pakaian jadi yang meliputi t-shirt,kemeja, tas yang memasarkan produknya pada *outlet-outlet* yang tersebar di beberapa wilayah Indonesia. Tugas utama *outlet* Blackjack yaitu melayani *outlet* secara langsung, sehingga keberadaan dari *outlet* ini sangatlah penting untuk dapat menentukan strategi bisnis Blackjack Industries ke depannya.

Setelah melakukan wawancara dengan pihak *marketting* Blackjack Industries, didapatkan informasi bahwa Blackjack Industries mengirimkan produk baru setiap sebulan sekali ke setiap *outlet* dan *outlet* menerima produk tersebut lalu memasarkan kepada *outlet*. Semua *outlet* akan menerima produk dan jumlah yang sama setiap pengiriman dilakukan. Berdasarkan survey kuesioner yang telah dilakukan terhadap 20 *outlet* Blackjack Industries diketahui bahwa 14 *outlet* mengalami kendala dalam hal ketidaksesuaian produk. Yang dimaksud dengan ketidaksesuaian produk adalah produk yang ada di *outlet* tidak sesuai dengan yang dibutuhkan *outlet* seperti kurang diminatinya produk yang di jual tiap *outlet* yang dilihat dari jenis dan model yang berakibat banyaknya produk yang tidak dibeli oleh *outlet*. Hal ini menyebabkan banyak produk yang tidak laku terjual. Kesimpulan yang di dapat bahwa produk yang dibutuhkan tiap *outlet* berbeda.

Berdasarkan masalah yang telah dipaparkan sebelumnya, solusi yang mungkin dilakukan adalah dengan membentuk segmentasi *outlet* sehingga karakteristik kebutuhan setiap *outlet* akan dapat terlihat. Kebutuhan dari setiap *outlet* dapat terlihat berdasarkan data penjualan yang telah dilakukan di masing-masing *outlet*. Tetapi hanya dengan melihat data penjualan di rasa kurang jelas dalam melihat karakteristik produk tiap *outlet*. Untuk memperjelas karakteristik produk tiap *outlet* dibutuhkan data penjualan, data produk dan data *outlet*. Dengan melihat dari data tersebut akan dapat menjadi pengetahuan untuk pihak *marketting* dalam menganalisis karakteristik atau perilaku yang berbeda yang mungkin membutuhkan produk atau pemasaran yang berbeda dari setiap segmen *outlet*. Sehingga diharapkan dapat memaksimalkan proses pemasaran produk yang tepat terhadap masing-masing *outlet*.

Teknik *clustering* dapat digunakan untuk mensegmentasikan *outlet* berdasarkan kedekatan karakteristik antar *outlet* [1]. Salah satu metode yang dapat digunakan pada teknik *clustering* adalah *Hierarchical Agglomerative Clustering* [8]. Metode tersebut akan digunakan pada penelitian ini karena dengan mengacu dari referensi yang ada metode tersebut akan melakukan pengelompokkan terhadap *cluster* atau objek yang belum diketahui. Selain itu, *cluster outlet* akan terbentuk secara alami berdasarkan karakteristik produk.

Oleh karena itu dari pertimbangan dan penjelasan yang sudah di paparkan di atas peneliti tertarik untuk melakukan penelitian di Blackjack Industries.

Berdasarkan pembahasan di atas maka maksud dari penelitian ini adalah untuk menerapkan *data mining* pada data penjualan, data produk dan data *outlet* produk untuk membentuk segmentasi *outlet* di Blackjack Industries.

Sedangkan tujuan yang ingin dicapai untuk penelitian ini adalah

1. Untuk membantu pihak Blackjak Industries dalam menentukan kebutuhan tiap *outlet* berdasarkan karakteristik yang di dapat.
2. Mempermudah *manager marketting* dalam menentukan rekomendasi produk yang tepat.

.

2. ISI PENELITIAN

2.1 Landasan Teori

Landasan teori pada penulisan skripsi ini akan menerangkan mengenai teori-teori yang berhubungan dengan aplikasi segmentasi *outlet* Blackjack Industries.

2.1.1 Data

Data adalah representasi fakta dunia nyata yang mewakili suatu objek seperti manusia (pegawai, siswa, pembeli, pelanggan), barang, hewan, peristiwa, konsep, keadaan, dan sebagainya, yang direkam dalam bentuk angka, huruf, simbol, teks, gambar, bunyi, atau kombinasinya[5].

2.1.2 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*[5].

2.1.3 CRISP-DM

CRISP – DM (*Cross – Industry Standard of Data Mining*). CRISP – DM merupakan metodologi data *mining* yang digambarkan dengan model proses dan di dalamnya terdapat tugas-tugas yang harus dilakukan untuk melakukan proyek data *mining*[2]

2.1.4 Clustering

Clustering merupakan pekerjaan yang memisahkan data / *vector* ke dalam sejumlah kelompok (*cluster*) menurut karakteristiknya masing – masing. Data – data yang memiliki kemiripan karakteristik yang berkumpul dalam *cluster* yang sama, dan data – data dengan karakteristik berbeda akan terpisah dalam *cluster* yang berbeda[4].

Menghitung jarak dengan menggunakan Euclidean :

$$d = \sqrt{\sum_{i=1}^{r}(X\,i - Yi)^2} \qquad \text{Persamaan 1}$$

Dimana :

d = Jarak

r = jumlah fitur dalam vektor

Xi = nilai fitur X ke i

Yi = nilai fitur Y ke i

2.1.5 Agglomerative Hierarchical Clustering (AHC)

Agglomerative hierarchical clustering merupakan algoritma yang digunakan untuk mengukur kedekatan diantara dua *cluster*. Terdapat tiga cara yang digunakan AHC, diantaranya adalah, metode *single linkage*, *complete linkage*, *average linkage*. [9].

Berikut ini merupakan persamaan *Maksimum linkage* :

$$d(U,V)=\text{maks}\{d(U,V)\};d(U,V)\in D \qquad \text{persamaan 2}$$

Berikut ini merupakan tahapan dari algoritma *Agglomerative* :

1. Merepresentasikan setiap outlet sebagai satu kelompok, kemudian menghitung jarak (*distance measure*) antar kelompok-kelompok tersebut.
2. Cari dua *cluster* yang mempunyai jarak paling dekat dan pilih nilai terbesar untuk menentukan nilai baru dari dua cluster yang terbentuk dengan menggunakan metode *complete linkage* yang terdapat pada persamaan (2).
3. Ulangi langkah kedua sampai tersisa satu kelompok (kumpulan semua outlet) akan berhenti ketika jumlah *cluster* yang terbentuk sama dengan 1.
4. Ketika Jumlah *cluster* berjumlah 1 maka tahap ke 3 telah selesai dan Hasil akhir dari perulangan berdasarkan ukuran dapat dilihat pada lampiran. Berikut ini merupakan dendrogram dari hasil pengelompokan

2.1.6 Agglomeration Schedule

Agglomeration schedule adalah ringkasan numerik dari solusi *cluster* yang dihasilkan dari proses *clustering* menggunakan metode hierarki [12].

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.

2.2 Analisis Masalah

Berdasarkan hasil wawancara yang telah dilakukan dengan pihak *marketing* maka didapatkan hasil sebagai berikut :

1. Terdapatnya kendala dalam hal ketidaksesuaian produk dari penjualan produk kepada pelanggan.
2. Tidak terdapatnya penentuan rekomendasi produk yang tepat.

Berdasarkan analisis masalah yang telah dipaparkan metode yang tepat untuk penyelesaian kasus yaitu dengan menggunakan metode *Data mining*. Dalam kasus ini akan menggunakan model proses CRISP-DM. Tahapan-tahapan model proses CRISP-DM sebagai berikut :

1. Pemahaman Proses Bisnis

Tahap pertama pada model proses CRISP-DM adalah tahap pemahaman bisnis (*business understanding*). Pada tahap ini, dilakukan pengumpulan informasi mengenai tujuan bisnis dan hasil yang diharapkan oleh manajer *marketing* Blackjack Industries. Kegiatan pada tahap ini meliputi identifikasi tujuan bisnis, penentuan tujuan *Data mining*.

Tujuan bisnis dari Blackjack Industries adalah meningkatkan kualitas produk di setiap *Outlet*. Hal tersebut dilakukan untuk mengetahui karakteristik untuk setiap produk, terutama jenis produk yang sesuai dengan kebutuhan pelanggan di setiap *Outlet*. Dengan melakukan hal ini diharapkan pelanggan akan melakukan pembelian kembali dan membeli produk lebih banyak lagi.

Kriteria Sukses Bisnis Menurut manajer *marketing*, aspek penting yang menjadi kriteria tercapainya tujuan bisnis di atas adalah dengan cara menjual produk yang tepat sesuai dengan karakteristik dari setiap *Outlet* kepada pelanggan, untuk melakukan hal tersebut maka Blackjack Industries perlu melakukan analisa ke setiap *Outlet* dan menentukan produk yang sesuai untuk setiap *Outlet*. Analisa yang dilakukan dengan menggunakan *Data* produk, *Outlet* dan penjualan sehingga pihak marketting dapat menentukan strategi bisnis yang tepat untuk ke depannya.

1.1.1 Penentuan Sasaran Data Mining

Kegiatan ini bertujuan untuk mendeskripsikan hasil yang diharapkan dari penelitian, sehingga tujuan bisnis dapat tercapai. Pada tahap ini dijelaskan tujuan proyek dari sudut pandang teknik. Berikut merupakan hasil dari penentuan sasaran *Data mining*.

a. Tujuan *Data* mining

Tujuan dari penerapan *Data* mining pada penjualan produk dalam penelitian ini adalah untuk membentuk segmentasi *Data Outlet* berdasarkan penjualan produk yang telah dilakukan oleh *Outlet.* Dengan penerapan *Data* mining ini diharapkan dapat membantu pihak *marketing* dalam menentukan strategi bisnis perusahaan.

b. Kriteria sukses *Data* mining

Kriteria sukses terhadap penelitian ini adalah dapat membantu pihak *marketing* Blackjack Industries dalam mengelompokan setiap *Outlet* yang memiliki karakteristik yang sama dalam penjualan produk.

2.2 Pemahaman *Data*

Tahapan pemahaman *Data* merupakan tahapan kedua yang dilakukan setelah tahapan pemahaman bisnis. Dalam tahapan pemahaman *Data* ini terdapat beberapa langkah diantaranya

2.2.2 Penjelasan *Data*

Data hasil join memiliki keseluruhan 10 atribut, 2 atribut mengenai *Outlet*, 5 atribut mengenai penjelasan jenis dan motif dari produk yang terjual.

2..2.3 Persiapan *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* hasil join yang telah mengalami proses pemilihan atribut. *Data* hasil pemilihan atribut merupakan *Data* set yang dapat dilihat pada Tabel 1.

Tabel 1. Persiapan Data Awal

ID Outlet	TS_OBLG	TS_RGLN	KMJ_PLS	KMJ_FLN	JKT_HOD	JKT_PRK	SWT_CRDG	SWT_RJT	POLO_PLS	BAG_PLS
1	9	8	4	11	4	5	4	2	8	11
2	13	11	8	11	7	4	5	4	10	7
3	7	13	2	15	5	7	3	5	9	8
4	15	8	7	12	5	5	3	4	8	6
5	14	12	3	8	3	6	5	1	10	8
6	9	11	5	9	6	8	4	4	9	3
7	14	13	8	9	5	3	7	7	11	13
8	11	7	3	14	6	6	3	6	9	8
9	10	13	5	11	4	6	5	2	9	9
10	13	11	8	13	5	6	5	5	9	9
11	13	14	6	13	6	5	5	5	9	11
12	10	13	7	12	10	4	6	3	9	9
13	14	11	5	12	9	7	4	3	8	7
14	14	8	6	8	1	8	3	5	8	7
15	10	10	6	8	7	3	4	5	11	9
16	11	12	4	13	6	5	6	3	11	9
17	7	9	5	11	4	6	5	2	9	9
18	13	13	8	11	7	4	5	4	10	7
19	11	9	2	15	5	7	3	5	9	8
20	11	8	3	14	6	6	3	6	9	8

Deskripsi *Data* Set

Pada deskripsi *Data* akan di jelaskan penamaan atribute.Untuk penjelasan dari nama atribut di *Data* set akan di jelaskan di bawah ini :

Keterangan nama tabel Hasil Pemilihan Atribut Dari *Data* Hasil Join (2):

TS = Jenis T-shirt
KMJ = Jenis kemeja
JKT = Jenis Jaket
SWT = Jenis Sweater
POLO = Jenis baju polo
BAG = Jenis tas
OBLG = Model Oblong
RGLN = Model Raglan
PLS = Model Polos
FLN = Model Flanner
HOD = Model Hoodie
PRK = Model Parka
CRD = Model Cardigan
RJT = Model Raju

2.4 Tahap Pemodelan

Tahapan selanjutnya dari kerangka kerja CRISP-DM adalah tahapan pemodelan, pada tahapan ini memilih dan mempersiapkan pemodelan yang sesuai dengan tujuan yang di capai.

2.4.1 Pemilihan Metode Pemodelan

Langkah pertama pada tahap pemodelan adalah memilih metode yang akan digunakan. Pemilihan metode disesuaikan dengan *Data* yang telah disiapkan. Metode yang akan digunakan adalah metode *hierarchical agglomerative clustering.*

Setelah metode pemodelan ditentukan, kemudian akan dijelaskan prosedur dari metode tersebut. Langkah-langkah dari metode *Agglomerative Hierarchical clustering* (AHC) adalah sebagai berikut.

1. Merepresentasikan setiap *Outlet* sebagai satu segmentasi, kemudian menghitung jarak (*distance measure*) antar segmentasi-segmentasi tersebut.
2. Cari dua *cluster* yang mempunyai jarak paling dekat dan pilih nilai terbesar untuk menentukan nilai baru dari dua *cluster* yang terbentuk dengan menggunakan metode *complete linkage* yang terdapat pada persamaan 2.
3. Ulangi langkah kedua sampai tersisa satu segmentasi (kumpulan semua *Outlet*) akan berhenti ketika jumlah *cluster* yang terbentuk sama dengan 1.
4. Ketika Jumlah *cluster* berjumlah 1 maka tahap ke 3 telah selesai dan Hasil akhir dari perulangan berdasarkan ukuran dapat dilihat pada lampiran.
5. 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.*

Langkah ke-1

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* (persamaan 1). setiap *cluster* akan dihitung jarak dengan dirinya sendiri dan jarak dengan *cluster* lainnya. Berikut ini merupakan proses perhitungan jarak antar *cluster* dengan menggunakan persamaan *Euclidean Distance* (persamaan 1). Berikut ini contoh perhitungan jarak berdasarkan ukuran :

Dman(1,2)=

$$\sqrt{(9-13)^2+(8-11)^2+(4-8)^2+\cdots+(11-7)^2}$$

= 8,72

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.

Langkah ke-2

Cari dua *cluster* yang mempunyai jarak paling dekat

Pada tahapan ini akan dihitung jarak antara dua *cluster* dengan menggunakan *complete linkage* (persamaan 2), dengan metode ini akan mencari jarak yang paling dekat dari dua *cluster* dan menggabung cluster dengan jarak maksimum dan menghasilkan cluster baru.

Berikut contohnya :

Tabel 3.10 Matriks *Euclidean* Distance

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	8,72	8,60	8,77	8,25	9,80	11,05	6,86	5,83	7,87	8,66	9,11	9,06	9,06	7,28	6,78	3,61	9,59	7,07	6,78
2	8,72	0,00	10,58	5,39	8,12	8,00	8,00	8,31	6,63	4,24	6,08	5,57	5,48	9,17	5,57	5,83	8,43	2,00	8,94	7,87
3	8,60	10,58	0	11,53	11,22	9,06	13,64	7,55	7,07	9,27	8,66	9,54	9,70	12,57	10,44	6,93	7,55	10,39	5,66	6,78
4	8,77	5,39	11,53	0	8,66	8,89	11,09	6,93	8,66	5,57	8,60	9,80	6,08	6,71	8,60	8,43	9,49	6,71	7,81	6,86
5	8,25	8,12	11,22	8,66	0	8,94	10,30	10,44	5,83	8,60	8,66	10,44	8,25	7,62	8,54	7,21	8,66	8,12	9,70	10,00
6	9,80	8,00	9,06	8,89	8,94	0	13,64	9,11	7,62	9,17	10,82	9,54	7,87	8,94	8,43	8,72	7,81	8,25	9,06	8,72
7	11,05	8,00	13,64	11,09	10,30	13,64	0	12,29	9,38	7,62	6,56	9,43	11,14	11,66	7,94	8,94	11,70	7,75	12,73	11,83
8	6,86	8,31	7,55	6,93	10,44	9,11	12,29	0	8,66	7,28	8,83	9,80	7,42	9,33	8,49	7,14	7,62	9,43	3,00	1,00
9	5,83	6,63	7,07	8,66	5,83	7,62	9,38	8,66	0	6,00	5,74	6,86	7,35	9,06	7,14	4,24	5,00	6,32	7,62	8,00
10	7,87	4,24	9,27	5,57	8,60	9,17	7,62	7,28	6,00	0	4,36	7,00	6,16	8,25	7,55	5,66	7,94	4,69	7,35	6,78
11	8,66	6,08	8,66	8,60	8,66	10,82	6,56	8,83	5,74	4,36	0	6,16	6,86	10,82	8,00	5,00	9,17	5,39	8,19	8,06
12	9,11	5,57	9,54	9,80	10,44	9,54	9,43	9,80	6,86	7,00	6,16	0	6,56	13,15	6,93	5,74	8,49	5,20	9,95	9,22
13	9,06	5,48	9,70	6,08	8,25	7,87	11,14	7,42	7,35	6,16	6,86	6,56	0	9,75	10,68	9,54	10,00	9,70	9,27	6,86
14	9,06	9,17	12,57	6,71	7,62	8,94	11,66	9,33	9,06	8,25	10,82	13,15	9,75	0	7,21	6,24	9,64	8,19	6,40	5,83
15	7,28	5,57	10,44	8,60	8,54	8,43	7,94	8,49	7,14	7,55	8,00	6,93	10,68	7,21	0	6,86	7,21	6,24	9,64	8,19
16	6,78	5,83	6,93	8,43	7,21	8,72	8,94	7,14	4,24	5,66	5,00	5,74	9,54	6,24	6,86	0	6,40	5,83	6,32	6,48
17	3,61	8,43	7,55	9,49	8,66	7,81	11,70	7,62	5,00	7,94	9,17	8,49	10,00	9,64	7,21	6,40	0	9,11	7,55	7,42
18	9,59	2,00	10,39	6,71	8,12	8,25	7,75	9,43	6,32	4,69	5,39	5,20	9,70	8,19	6,24	5,83	9,11	0	9,59	8,83
19	7,07	8,94	5,66	7,81	9,70	9,06	12,73	3,00	7,62	7,35	8,19	9,95	9,27	6,40	9,64	6,32	7,55	9,59	0	2,45
20	6,78	7,87	6,78	6,86	10,00	8,72	11,83	1,00	8,00	6,78	8,06	9,22	6,86	5,83	8,19	6,48	7,42	8,83	2,45	0

(dman)=(d(8,20))=1. Dari semua *cluster* yang ada akan dipilih dua *cluster* yang berbeda seperti *cluster* 1 dan 2 dengan melihat dari nilai paling minimum dari semua *cluster*. Dua *cluster* yang sama seperti *cluster* 1 dan 1 atau 2 dan 2 tidak dipilih karena merupakan *cluster* yang sama dan memiliki nilai minimum yaitu 0. Untuk awal pemilihan *cluster* tepilih *cluster* 8 dan 20 karena memiliki nilai paling minimum dari *cluster* lain. Menghitung jarak 8 dan 20 dengan *cluster* lain yang tersisa. Contoh sebagai berikut:

d(8,20) 4= maks {d(8,4) d(20,4)}= maks {(6,93) (6,86)}= 6,93 karena setelah membandingkan 2 nilai

yaitu 6,93 dan 6,86 dipilih nilai yang paling maksimum. Setelah melakukan perhitungan dengan segmentasi yang lain Cluster d(8,20) akan membentuk cluster baru dan cluster 8 dan 20 akan di hapus. Dan akan muncul cluster baru yaitu 8,20. dalam pemilihan nilai paling maksimum ini dilakukan untuk mendapatkan nilai baru untuk cluster baru. Contoh sebagai berikut:

Tabel 3.11 Matriks *Euclidean* Distance Iterasi-1

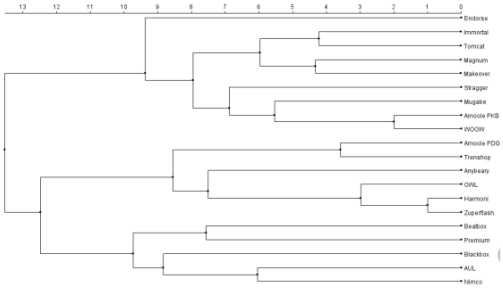
	1	2	3	4	5	6	7	8,20	9	10	11	12	13	14	15	16	17	18	19
1	0	8,72	8,60	8,77	8,25	9,80	11,05	6,86	5,83	7,87	8,66	9,11	9,06	9,06	7,28	6,78	3,61	9,59	7,07
2	8,72	0,00	10,58	5,39	8,12	8,00	8,00	8,31	6,63	4,24	6,08	5,57	5,48	9,17	5,57	5,83	8,43	2,00	8,94
3	8,60	10,58	0	11,53	11,22	9,06	13,64	7,55	7,07	9,27	8,66	9,54	9,70	12,57	10,44	6,93	7,55	10,39	5,66
4	8,77	5,39	11,53	0	8,66	8,89	11,09	6,93	8,66	5,57	8,60	9,80	6,08	6,71	8,60	8,43	9,49	6,71	7,81
5	8,25	8,12	11,22	8,66	0	8,94	10,30	10,44	5,83	8,60	8,66	10,44	8,25	7,62	8,54	7,21	8,66	8,12	9,70
6	9,80	8,00	9,06	8,89	8,94	0	13,64	9,11	7,62	9,17	10,82	9,54	7,87	8,94	8,43	8,72	7,81	8,25	9,06
7	11,05	8,00	13,64	11,09	10,30	13,64	0	12,29	9,38	7,62	6,56	9,43	11,14	11,66	7,94	8,94	11,70	7,75	12,73
8,20	6,86	8,31	7,55	6,93	10,44	9,11	12,29	0	8,66	7,28	8,83	9,80	7,42	9,33	8,49	7,14	7,62	9,43	3,00
9	5,83	6,63	7,07	8,66	5,83	7,62	9,38	8,66	0	6,00	5,74	6,86	7,35	9,06	7,14	4,24	5,00	6,32	7,62
10	7,87	4,24	9,27	5,57	8,60	9,17	7,62	7,28	6,00	0	4,36	7,00	6,16	8,25	7,55	5,66	7,94	4,69	7,35
11	8,66	6,08	8,66	8,60	8,66	10,82	6,56	8,83	5,74	4,36	0	6,16	6,86	10,82	8,00	5,00	9,17	5,39	8,19
12	9,11	5,57	9,54	9,80	10,44	9,54	9,43	9,80	6,86	7,00	6,16	0	6,56	13,15	6,93	5,74	8,49	5,20	9,95
13	9,06	5,48	9,70	6,08	8,25	7,87	11,14	7,42	7,35	6,16	6,86	6,56	0	9,75	10,68	9,54	10,03	9,70	9,27
14	9,06	9,17	12,57	6,71	7,62	8,94	11,66	9,33	9,06	8,25	10,82	13,15	9,75	0	7,21	6,24	9,64	8,19	6,40
15	7,28	5,57	10,44	8,60	8,54	8,43	7,94	8,49	7,14	7,55	8,00	6,93	10,68	7,21	0	6,86	7,21	6,24	9,64
16	6,78	5,83	6,93	8,43	7,21	8,72	8,94	7,14	4,24	5,66	5,00	5,74	9,54	6,24	6,86	0	6,40	5,83	6,32
17	3,61	8,43	7,55	9,49	8,66	7,81	11,70	7,62	5,00	7,94	9,17	8,49	10,00	9,64	7,21	6,40	0	9,11	7,55
18	9,59	2,00	10,39	6,71	8,12	8,25	7,75	9,43	6,32	4,69	5,39	5,20	9,70	8,19	6,24	5,83	9,11	0	9,59
19	7,07	8,94	5,66	7,81	9,70	9,06	12,73	3,00	7,62	7,35	8,19	9,95	9,27	6,40	9,64	6,32	7,55	9,59	0

Langkah ke-3

Ulangi langkah kedua sampai tersisa satu segmentasi (kumpulan semua *Outlet*) langkah akan berhenti ketika jumlah *cluster* yang terbentuk sama dengan 1.

Langkah ke-4

Ketika Jumlah *cluster* berjumlah 1 maka tahap ke 3 telah selesai dan Hasil akhir dari perulangan berdasarkan ukuran dapat dilihat pada lampiran. Berikut ini merupakan dendrogram dari hasil segmentasi. Hasil dendgoram pada Gambar 3.1 Hasil Akhir Dendogram seperti berikut :



Gambar 3.1 Hasil Akhir Dendogram

Langkah ke-5

Menghasilkan *cluster* dengan memotong *dendrogram* pada tingkat yang tepat

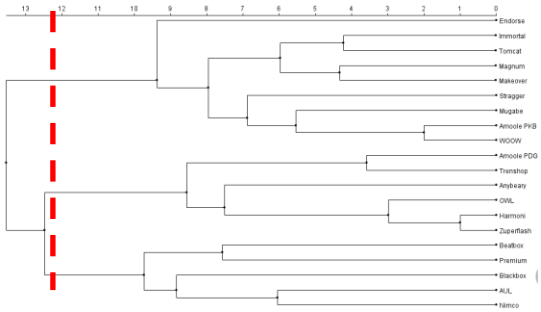
Berdasarkan dendrogram yang telah dibentuk, tingkat yang tepat atau jumlah *cluster* terbaik akan ditentukan dengan melihat selisih jarak dari setiap *cluster* yang terbentuk. Untuk mengetahui jumlah *cluster* terbaik salah satu caranya dapat menggunakan *agglomeration schedule. Agglomeration schedule* dapat dilihat pada Tabel 3.12.

Jumlah *cluster* yang terbaik dapat dilihat dari selisih koofisien dari setiap iterasinya. Selisih dari koofisien dapat dilihat pada Tabel 3.13.

Tabel 3.13 Selisih Koofisien Segmentasi Berdasarkan Ukuran

Jumlah Cluster	Tahap Terakhir Agglomeration	Koefisien Tahap ini	Selisih
2	13,64	12,57	1,07
3	12,57	9,75	2,82
4	9,75	9,43	0,32
5	9,43	8,89	0,55
6	8,89	8,60	0,29
7	8,60	8,00	0,60
8	8,00	7,62	0,38
9	7,62	7,55	0,07
10	7,55	6,93	0,62
11	6,93	6,08	0,85
12	6,08	6,00	0,08
13	6,00	5,57	0,43
14	5,57	4,36	1,21
15	4,36	4,24	0,12
16	4,24	3,61	0,64
17	3,61	3,00	0,61
18	3,00	2	1,00
19	2	1	1

Selisih terbesar yaitu 2.82 sehingga jumlah *cluster* terbaik adalah 3 *cluster*. Maka, dendrogram hasil segmentasi berdasarkan ukuran akan dipotong seperti terlihat pada Gambar 3.2



Tabel 3.12 *Agglomerative Schedule*

Iterasi	Kombinasi Cluster		Koefisien	Iterasi Kombinasi Cluster sebel	
	Cluster 1	Cluster 2		Cluster 1	Cluste
1	8	20	1	0	
2	2	18	2,00	0	
3	8	19	3,00	1	
4	1	17	3,61	0	
5	9	16	4,24	0	
6	10	11	4,36	0	
7	2	12	5,57	2	
8	9	10	6,00	5	
9	4	13	6,08	0	
10	2	15	6,93	7	
11	3	8	7,55	0	
12	5	14	7,62	0	
13	2	9	8,00	10	
14	1	3	8,60	4	
15	4	6	8,89	9	
16	2	7	9,43	13	
17	4	5	9,75	15	
18	1	4	12,57	14	
19	1	2	13,64	18	

Dari hasil segmentasi *Outlet* berdasarkan penjualan produk didapatkan hasil tertinggi dari setiap *cluster* yang digunakan sebagai acuan untuk penentuan karakteristik tiap *cluster*, sebagai berikut :

Tabel 3.20 Penjualan Tertinggi Pada Setiap *Cluster*

Cluster	Jenis dan Model	Jumlah
1	KMJ_FLN	15
2	TS-OBLG	14
	TS-RGLN	14
3	TS-OBLG	15

2.5 Evaluasi

Pada tahap ini, model sudah terbentuk dan diharapkan memiliki kualitas baik jika dilihat dari sudut pandang analisa *Data*. Pada tahap ini akan dilakukan evaluasi terhadap keefektifan dan kualitas model sebelum digunakan dan menentukan apakah model dapat mencapat tujuan yang ditetapkan pada fase awal (*Business Understanding*).

2.5.1 Evaluasi Hasil

langkah ini menilai sejauh mana model memenuhi tujuan bisnis dan berusaha untuk menentukan apakah ada beberapa alasan bisnis mengapa model ini kekurangan.

2.5.2 Penilaian *Data* mining hasil sehubungan dengan kriteria keberhasilan bisnis

Penilaian yang dilakukan ialah dengan menghubungkan kriteria keberhasilan bisnis dengan model yang sudah terbentuk. Kriteria keberhasilan bisnis yang diharapkan adalah dengan menjual produk yang tepat sesuai dengan karaketeristik dari setiap *Outlet*. Sedangkan model yang sudah terbentuk dapat dilihat pada hasil rekomendasi produk.

2.5.3 Model disetujui

Model yang disetujui dapat memenuhi kriteria keberhasilan bisnis. Dengan melihat dari penilaian data mining terhapa kriteria keberhasilan bisnis maka kesimpulan yang didapat adalah model yang sudah terbentuk yang memiliki karakteristik sudah memenuhi kriteria keberhasilan bisnis. Dengan melihat hasil ini maka model yang terbentuk dapat disetujui dan dapat digunakan sebagai acuan untuk pihak *marketting* dalam penentuan strategi ke depannya.

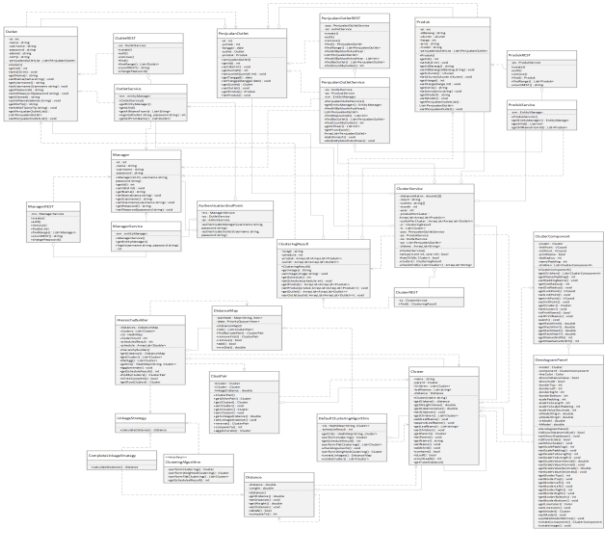
2.6 Use Case Diagram

Use Case Diagram digunakan untuk menggambarkan kebutuhan fungsional dari sistem yang akan dibangun. Use Case Diagram menjelaskan hubungan antara pengguna dengan sistem. Diagram ini menjadi titik awal bagi pengembangan sistem menggunakan pendekatan berorientasi objek. Gambar 3.3 merupakan Use Case Diagram sistem segmentasi Outlet pada penelitian ini.

,

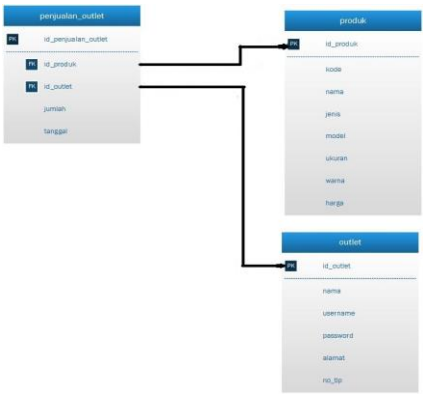
2.7 Class Diagram

Diagram atau *class diagram* menggambarkan struktur sistem dari segi pendefinisian kelas-kelas yang akan dibuat untuk membangun sistem. Kelas memiliki apa yang disebut atribut dan metode atau operasi. *Class Diagram* yang dibutuhkan untuk membangun sistem.



2.8 Skema Relasi

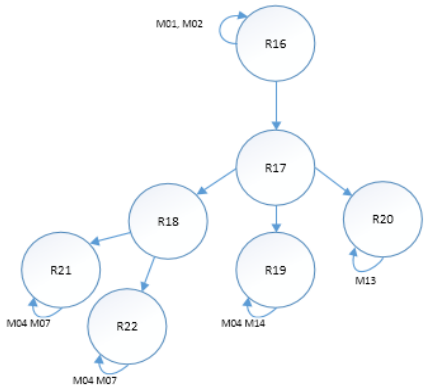
Skema relasi merupakan rangkaian hubungan antara dua tabel atau lebih pada sistem basis *Data*. Berikut ini adalah skema relasi pada sistem pembentukan segmentasi *Outlet* pada Blacjack Industries :



2.9 Jaringan Semantik

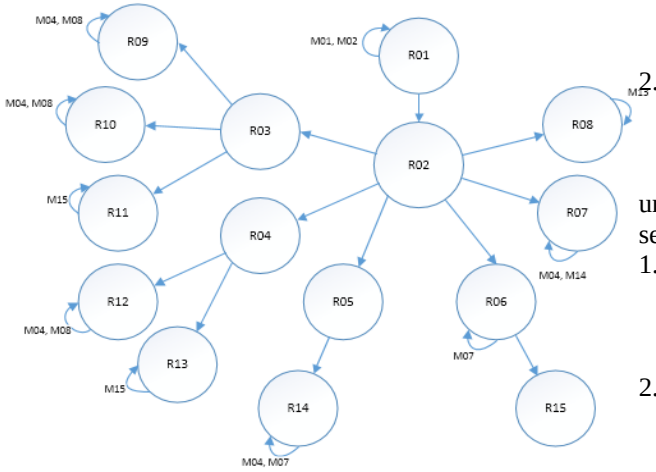
Jaringan semantik menggambarkan hubungan antara komponen-komponen antarmuka yang terdiri *form* dan pesan. Gambar 3.56 merupakan jaringan semantik dari sistem segmentasi *Outlet*.

1. Jaringan Semantik *Outlet*



Gambar 3.56 Jaringan Semantik *Outlet*

2. Jaringan Semantik *Manajer Marketting*



Gambar 3.57 Jaringan Semantik *Manager Marketting*

2.10 Pengujian

Pengujian aplikasi dilakukan terhadap fungsionalitas aplikasi untuk menilai apakah aplikasi yang dibangun sudah memenuhi kriteria atau terdapat ketidaksesuaian pada fungsionalitas sebagaimana yang terdapat pada perancangan aplikasi. Pengujian yang dilakukan untuk menguji aplikasi adalah pengujian black box.

Rencana pengujian yang akan dilakukan adalah dengan cara menguji aplikasi yang dibangun secara black box dan beta. Pengujian black box terfokus pada pengujian fungsionalitas aplikasi yang dibangun.

2.10.1 Skenario Black Box

Pengujian black box difokuskan pada persyaratan fungsionalitas perangkat lunak yang akan dibangun. Rencana pengujian selengkapnya dapat dilihat pada Tabel 4.9.

Tabel 4.9 Rencana Pengujian Aplikasi

Kelas Uji
Login pengguna
Edit data produk
Hapus data produk
Tambah data produk
Hapus data outlet
Tambah data outlet
Edit data penjualan
Hapus data penjualan
Tambah data penjualan
Tambah data penjualan dengan excel
Lihat data penjualan
Melakukan *Clustering*

3. PENUTUP

Dari hasil penelitian ini dapat kesimpulan bahwa

1. Aplikasi pembentukan segmentasi *outlet* yang telah dibangun dapat membantu *Manager Marketting* dalam menentukan segmentasi *outlet* yang sesuai dengan karakteristik dari penjualan produk sehingga dapat dijadikan acuan untuk strategi bisnis yang tepat ke depannya.
2. Mempermudah *Manager Marketting* dalam menentukan rekomendasi produk yang tepat.

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

1. Sistem dapat memberikan fasilitas untuk berkomunikasi antara *admin outlet* dengan pihak *manager marketing* yang berguna sebagai sarana informasi.
2. Dalam melakukan analisis data penjualan produk pihak *manager marketing* dapat memilih kategori produk seperti jenis, model, ukuran dan warna untuk dilakukan analisis, sehingga sistem dapat menentukan karakteristik *outlet* dari kategori produk yang dipilih.

DAFTAR PUSTAKA

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

[2] C. Pete (NCR), et al., "*CRISP-DM 1.0 Step-by-step data mining guide*", SPSS, 2000.

[3] R.S. Arini dan M. Shalahuddin, "Modul Pembelajaran Rekayasa Perangkat Lunak (Terstruktur dan Berorientasi Objek)", Bandung: Modula, 2013.

[4] P. Eko., Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab, Yogyakarta: Andi, 2014.

[5] J. Han dan M. Kamber, *Data Mining* Concepts and Techniques, Jerman: Morgan Kauffman, 2006.

[6] A. Edgar, and C. Rodriguez. "The treatment of missing values and its effect on classifier accuracy." *Classification, clustering, and data mining applications*. Springer Berlin Heidelberg, 2004.

[7] Ir. Supardi, Y., "Semua Bisa Menjadi *Programming Java Basic Programming*", Jakarta: Elex Media, 2010.
[8] B. Susanto, "*Data Preprocessin*g,", Versi 1.2, Jakarta: Elex Media Komputindo ,2013.
[9] P. Berkhin, *Survey of Clustering Data Mining Techniques*, Jerman: Accrue Software, 2006.
[10] Hermawati, Fajar Astuti, *Data Mining*, Yogyakarta: Andi, 2013.
[11] Xu, Rui, and Don Wunsch, *Clustering*, New Jersey: John Wiley & Sons. Inc, 2008.
[12] Caccam, Marielle., and Refran Jewel., *Clustering analysis, 2012, Wired*, 06 Desember 2012, [Online]. Tersedia: http://www.slideshare.net/jewelmrefran/*cluster*-analysis-15529464 [Diakses: 05 Januari 2017].
[13] L. Sofya, "Analisis *Cluster* dengan *Average Linkage Method* dan *Ward's Method* untuk Data Responden Nasabah Asuransi Jiwa Unit Link", pp. 84-91, Des 2014.