

Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan

By Wawan LY Saptomo

Penerapan *Agglomerative Hierarchical Clustering* Untuk Segmentasi Pelanggan

Widya Gati¹⁾, Wawan Laksito Yuly Saptomo²⁾, Yustina Retno Wahyu Utami³⁾

¹⁾ Program Studi Sistem Informasi, STMIK Sinar Nusantara Surakarta

^{2,3)} Program Studi Teknik Informatika, STMIK Sinar Nusantara Surakarta

¹⁾widyaadelta88@gmail.com; ²⁾wlaksito@sinus.ac.id; ³⁾yustina.retno@gmail.com

ABSTRACT

As more businesses emerge, companies need to have the right marketing strategy to provide the best service to customers. The first step is to know the type of customer and make appropriate marketing strategies according to the type of customer. In this research, it is proposed for clustering customers so that an appropriate strategy for that customer group can be determined. The method used for cluster formation uses Agglomerative Hierarchical Clustering with Average Linkage approach and distance determination using Manhattan Distance. The variables in this research are Recency, Frequency, and Monetary (RFM). The results of testing using the Silhouette coefficient show that the results of 7 clusters are the best results when compared with 2 clusters up to 20 clusters because they have the smallest minus value. Based on the results of the Silhouette coefficient, customer segmentation uses 7 clusters with each cluster representing the existing customer type.

Keywords: *Customer Segmentation, Agglomerative Hierarchical Clustering, Average Linkage, Clustering*

I. PENDAHULUAN

Seiring dengan banyaknya toko komputer lain yang bermunculan, perusahaan ingin memberikan pelayanan yang terbaik agar dapat bersaing dengan yang lain, namun sampai saat ini belum memiliki strategi pemasaran yang khusus dan belum tepat sasaran. Oleh karena itu, dirancang sistem *clustering* pelanggan atau *customer* untuk mengetahui tipe pelanggan, sehingga dapat membuat strategi pemasaran yang tepat sesuai dengan tipe pelanggan yang ada.

Penerapan metode *Agglomerative Hierarchical Clustering* (AHC) dapat membantu dalam pembuatan *clustering customer*. Metode AHC dipilih karena walaupun lebih lambat, hasil akhir pengelompokan *Agglomerative Hierarchical Clustering* lebih baik dibandingkan *K-Means* yang sering digunakan. Hal ini didasarkan pada penelitian sebelumnya yakni Perbandingan Kinerja Metode *Complete Linkage*, Metode *Average Linkage*, Dan Metode *K-Means* Dalam Menentukan Hasil Analisis *Cluster*. Dalam penelitian tersebut dijelaskan bahwa metode AHC merupakan metode yang lebih baik daripada *K-Means* terutama metode AHC *Complete Linkage* dan *Average Linkage*. Untuk menentukan *cluster* yang sesuai dengan pelanggan perusahaan, *customer* dinilai profitabilitasnya terhadap perusahaan dari transaksi yang telah dilakukan dengan melihat RFM (*Recency, Frequency, dan Monetary*). *Recency* merupakan lamanya interval waktu sejak pelanggan melakukan transaksi terakhir, *frequency* berkaitan dengan tingkat keseringan pelanggan melakukan transaksi, dan *Monetary* yaitu besarnya nilai transaksi yang dilakukan selama periode tertentu. Untuk penentuan strategi pemasaran, salah satu strategi pemasaran yang relevan untuk diterapkan adalah strategi *Customer Relationship Management* (CRM).

Penelitian yang sejenis sudah dilakukan oleh Gita Febrina Wulandari untuk Segmentasi Pelanggan Pada Hijab Miulan. Algoritma yang digunakan untuk pembentukan *cluster* adalah algoritma *K-Means* dan menggunakan atribut RFM (*Recency, Frequency dan Monetary*) dengan tools Rapidminer 5.3. Terbentuk empat cluster yang memenuhi

kelas pada Piramida Pelanggan yaitu *Most Valuable Customer* ada 4 pelanggan, *Most Growable Customer* ada 23 pelanggan, *Below Zero* 190 pelanggan dan *Migrators* ada 115 pelanggan. Strategi yang diterapkan pada *Most Valuable Customer* dan *Most Growable Customer* adalah potongan harga, sedangkan kategori *Below Zero* dan *Migrators* pemasaran melalui email, media sosial dan menggunakan web perusahaan untuk mencari calon pelanggan baru.

Tujuan dari penelitian ini adalah terciptanya sistem *Agglomerative Hierarchical Clustering* (AHC) untuk *clustering customer* dan membantu perusahaan dalam menentukan strategi pemasaran yang tepat sesuai dengan jenis *customer* terhadap profitabilitas perusahaan.

II. TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Ada beberapa penelitian terkait *clustering customer* sebelumnya dengan menggunakan bermacam – macam metode, namun belum ada yang menggunakan metode *Agglomerative Hierarchical Clustering*. Metode AHC dipilih karena dapat menangani data dalam jumlah banyak dan hasil *cluster* yang diberikan lebih baik daripada metode *K-Means*.

Penelitian pertama tentang Segmentasi Pelanggan Menggunakan Algoritma *K-Means* Untuk *Customer Relationship Management* (CRM) Pada Hijab Miulan. Penelitian tersebut dilakukan karena dirasa strategi pemasaran yang diterapkan belum maksimal. Algoritma yang digunakan untuk pembentukan cluster adalah algoritma *K-Means* dan menggunakan atribut RFM (*Recency, Frequency* dan *Monetary*) dengan tools Rapidminer 5.3. Terbentuk empat cluster yang memenuhi kelas pada Piramida Pelanggan yaitu *Most Valuable Customer* ada 4 pelanggan, *Most Growable Customer* ada 23 pelanggan, *Below Zero* 190 pelanggan dan *Migrators* ada 115 pelanggan. Strategi yang diterapkan pada *Most Valuable Customer* dan *Most Growable Customer* adalah potongan harga, sedangkan kategori *Below Zero* dan *Migrators* pemasaran melalui email, media sosial dan menggunakan web perusahaan untuk mencari calon pelanggan baru (Wulandari, 2014).

Penelitian kedua mengenai Model Segmentasi Pelanggan Dengan *Kernel K-Means Clustering* Berbasis *Customer Relationship Management*. Data yang digunakan dalam penelitian ini berasal dari database Kantor Pos Medan tahun 2011 – bulan maret 2013. Dalam jurnal ini algoritma *Kernel K-Means Clustering* telah digunakan untuk mendapatkan suatu model aturan menentukan pelanggan terbaik dan potensial Kantor Pos Medan. Model aturan yang diperoleh menunjukkan bahwa katagori pelanggan terbaik dapat diperoleh jika transaksinya banyak dan besar uangnya sedang dan tinggi (Lubis, 2016).

Penelitian berikutnya bertema Implementasi Algoritma *Agglomerative Hierarchical Clustering* Untuk Mengelompokkan Capaian Belajar Siswa SD. Penelitian tersebut dilakukan karena proses pendampingan dan pembelajaran siswa kurang efektif, maka dari itu dilakukan pengelompokan untuk mempermudah guru dalam proses belajar mengajar. Pengelompokan menggunakan metode *Agglomerative Hierarchical Clustering* dengan menggunakan data nilai UTS dan UAS. Dari hasil pengelompokan dengan menggunakan metode *Single Linkage, Average Linkage* dan *Complete Linkage* didapatkan hasil pengelompokan yang paling seimbang pada metode *Average Linkage* dengan masing-masing jumlah data pada cluster 1, 2 dan 3 sebanyak 14, 22 dan 31 (Kurniawan, 2017).

Penelitian selanjutnya yakni Analisis *Cluster* dengan *Average Linkage Method* dan *Ward's Method* untuk Data Responden Nasabah Asuransi Jiwa Unit Link. Penelitian tersebut dilakukan untuk mengetahui langkah-langkah analisis *cluster* dengan metode *average linkage* dan metode *Ward*, serta membandingkan hasil analisis kedua metode

tersebut. Hasil pengclusteran kedua metode tersebut dibandingkan dengan kriteria simpangan baku dalam kelompok (SW) dan simpangan baku antar kelompok (SB). Metode terbaik memiliki nilai rasio SW dan SB yang lebih kecil. Hasil penelitian menunjukkan bahwa metode *average linkage* memiliki nilai rasio SB dan SW sebesar 0,486 sedangkan metode *Ward* memiliki nilai rasio SB dan SW sebesar 0,710. Hal tersebut menunjukkan bahwa metode *average linkage* memiliki kinerja lebih baik daripada metode *Ward* (Laeli, 2014).

Penelitian mengenai segmentasi konsumen produk hijau sayuran organik telah dilakukan. Atribut yang digunakan berdasarkan NMI (Natural Marketing Institute) yaitu LOHAS, Naturalites, Drifters, Conventionals dan Unconcerned. Metode pengambilan sampel yang digunakan adalah non-probability sampling, dengan metode purposive sampling sedangkan Metode analisis yang digunakan adalah deskriptif kuantitatif dan alat proses data yang digunakan adalah K-Means. Hasil penelitian ini diperoleh bahwa konsumen kategori drifters adalah konsumen paling dominan dengan persentase mencapai 43% dan memiliki karakteristik konsumen naturalites dan conventionals, konsumen kategori drifters memiliki pemahaman yang cukup baik dalam melindungi lingkungan. Konsumen kategori conventionals menempati urutan kedua dengan jumlah responden sebesar 27.5% dari total sampel. Konsumen conventionals cenderung berhemat dalam menggunakan energi namun dengan motivasi yang berbeda dengan yang dilakukan oleh konsumen kategori LOHAS ataupun naturalites. Konsumen kategori naturalites menempati urutan ketiga dengan persentase sebesar 22.5% dari total sampel. Konsumen kategori naturalites sangat aktif dalam menghemat energi dan memboikot perusahaan yang tidak ramah lingkungan, konsumen kategori ini cenderung memiliki gaya hidup sehat hanya untuk diri sendiri dan tidak terlibat apapun dalam konservasi lingkungan hidup. Konsumen kategori LOHAS menempati urutan keempat dengan persentase sebesar 4% dari total sampel. Karakteristik konsumen LOHAS melakukan boikot kepada perusahaan yang memiliki proses produksi tidak ramah lingkungan, dan sangat aktif dalam konservasi lingkungan hidup dan terlibat dalam kampanye hijau. Konsumen kategori unconcerned menempati urutan terakhir dengan persentase sebesar 3%. Karakteristik konsumen kategori unconcerned adalah mereka cenderung tidak peduli dengan keadaan lingkungan sekitar dan tidak ingin terlibat apapun dalam konservasi hijau (Effendi, 2015).

Penentuan strategi pemasaran promosi penjualan sepeda motor dilakukan dengan melakukan segmentasi pelanggan. Atribut yang digunakan yaitu jenis motor, jenis pembelian dan pekerjaan pada dealer Honda. Proses pengelompokan konsumen dilakukan dengan menggunakan Algoritma K-Means dengan membagi data dalam k *cluster*. Hasil perbandingan aplikasi dengan menggunakan software SPSS diperoleh hasil selisih dari masing-masing cluster dengan jumlah rata-rata 7%. Hal ini disebabkan karena centroid awal yang dipilih secara acak (Sulistiyowati, Ketherin, Arifiyanti, & Sodik, 2018).

2.2 Segmentasi Pelanggan

Segmentasi atau *Clustering* adalah metode untuk mengidentifikasi kelompok alami dari sebuah kasus yang didasarkan pada sebuah kelompok atribut. Segmentasi mengelompokkan data yang memiliki kemiripan atribut. *Clustering* bersifat *unsupervised*, karena tidak ada satu atribut pun yang digunakan untuk memandu proses pembelajaran, jadi seluruh atribut input diperlakukan sama. Sebagian besar algoritma *clustering* akan membangun sebuah model melalui serangkaian iterasi dan berhenti ketika model tersebut telah memusat atau berkumpul (batasan dari segmentasi telah stabil) (Vulandari, 2017).

Sedangkan pelanggan didefinisikan sebagai sebuah individu atau kelompok yang melakukan pembelian atas sebuah produk atau jasa berdasarkan keputusan akan pertimbangan harga dan penawaran yang berkomunikasi dengan perusahaan melalui surat, panggilan telepon, dan email yang dikirimkan secara berkala (Greenberg & Baron, 2010).

Proses segmentasi pelanggan akan membagi pelanggan menjadi beberapa kelompok yang disesuaikan dengan tingkat keloyalitasannya. Segmentasi pelanggan dibagi menjadi 6 karakteristik berdasarkan nilai RFM (Hardiani, Sulisty, & Hartanto, 2015) dapat dilihat pada Tabel 1.

Tabel 1. Karakteristik Pelanggan

Jenis Pelanggan	Karakteristik
<i>Superstar</i>	a. Mempunyai nilai <i>monetary</i> yang tinggi. b. Mempunyai <i>frequency</i> yang tinggi. c. Mempunyai <i>recency</i> yang tinggi.
<i>Golden customer</i>	a. Mempunyai nilai <i>monetary</i> tertinggi yang ke dua. b. <i>Frequency</i> yang tinggi. c. Mempunyai <i>recency</i> yang tinggi.
<i>Typical customer</i>	Mempunyai rata-rata nilai <i>monetary</i> dan rata-rata nilai <i>recency</i> .
<i>Occasional customer</i>	a. Nilai <i>monetary</i> tinggi b. Nilai <i>frequency</i> rendah. c. <i>Recency</i> rendah.
<i>Everyday shopper</i>	a. Memiliki <i>recency</i> yang tinggi. b. <i>Frequency</i> yang tinggi. c. Mempunyai nilai <i>monetary</i> rendah.
<i>Dormant customer</i>	a. Mempunyai <i>frequency</i> dan <i>monetary</i> yang paling rendah. b. Nilai <i>recency</i> yang paling rendah.

2.3 Customer Relationship Customer (CRM)

CRM adalah strategi organisasi yang fokus pada pelanggan. CRM meliputi semua aspek berhubungan dengan calon pelanggan dan pelanggan yang dimiliki oleh perusahaan saat ini. Lingkup CRM termasuk di dalamnya adalah pusat panggilan (*call center*), tenaga penjualan (*sales force*), pemasaran (*marketing*), dukungan teknis (*technical support*) dan layanan lapangan (*field service*) (Prasetyo, 2012).

CRM diawali dengan proses pemasaran dimana perusahaan menentukan target dan calon pelanggan yang berpotensi sebagai tujuan pemasaran tersebut (Rainer & Cegielski, 2011).

2.4 Agglomerative Hierarchical Clustering (AHC)

Clustering dengan pendekatan hirarki akan mengelompokkan data yang mirip dalam hirarki yang sama dan yang tidak mirip di hirarki yang agak jauh. Terdapat dua metode yang sering digunakan yaitu *agglomerative hierarchical clustering* dan *divisive hierarchical clustering*. *Agglomerative* melakukan *clustering* dari *N cluster* menjadi satu kesatuan *cluster*, dimana *N* adalah jumlah data, sebaliknya *divisive* melakukan proses *clustering* dari satu *cluster* menjadi *N cluster* (Everitt, Landau, Leese, & Stahl, 2011).

Beberapa metode *hierarchical clustering* yang sering digunakan dibedakan menurut cara perhitungan tingkat kemiripan. Ada yang menggunakan *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Average Group Linkage* dan lain-lain. Seperti juga halnya dengan *partition-based clustering*, jarak dapat digunakan untuk menghitung tingkat kemiripan antar data.

2.4.1 Single Linkage

Metode *Single Linkage clustering* (SL) juga disebut *nearest-neighbor technique* dimana pencarian pasangan berdasarkan pengukuran jarak terdekat. Sebut saja *G* dan *H* adalah dua cluster yang akan disatukan. Ketidaksamaan jarak $d(G,H)$ akan dihitung lalu dengan cara membandingkan setiap jarak anggota kelompok dari G_i terhadap jarak setiap anggota kelompok dari H_i kemudian mencari pasangan yang jaraknya terdekat.

$$d_{SL}(G, H) = \min(d_{i i'}); i \in G; i' \in H \dots\dots\dots(1)$$

2.4.2 Complete Linkage

Metode *Complete Linkage Agglomerative Clustering* (CL) disebut juga metode *furthest neighbor technique*. Tahapan metode ini secara umum hampir sama dengan metode *single linkage* tetapi pada pencarian pasangan, metode *complete linkage* mencari pasangan yang memiliki jarak terjauh dari nilai observasi.

$$d_{CL}(G, H) = \max(d_{i'i'}; i \in G; i' \in H) \dots\dots\dots(2)$$

2.4.3 Average Linkage

Selain metode *Single Linkage clustering* dan *Complete Linkage Agglomerative Clustering*, terdapat metode lain yaitu *Average Linkage* atau disebut juga *Group Average*(GA). Pada metode ini, pencarian pasangan ditentukan dengan melihat rata-rata jarak setiap nilai observasinya.

$$d_{GA}(G, H) = \frac{1}{N_G N_H} \sum_{i \in G} \sum_{i' \in H} d_{ii'} \dots\dots\dots(3)$$

Hierarchical clustering dapat digambarkan melalui dendogram. Dendogram disusun dengan membuat *similarity matrix* yang memuat tingkat kemiripan antar data yang dikelompokkan. Tingkat kemiripan bisa dihitung dengan berbagai macam cara seperti dengan *Euclidean Distance Space* dan *Manhattan Distance*. Perhitungan jarak dengan *Manhattan Distance* dan *Euclidean Distance* diperlihatkan pada rumus (4) dan (5).

a. *Manhattan Distance*

$$D_{man}(x, y) = \sum_{j=1}^d |x_j - y_j| \dots\dots\dots(4)$$

b. *Euclidean Distance*

$$D_{(x_2, x_1)} = \sqrt{\sum_{j=1}^d |x_{2j} - x_{1j}|^2} \dots\dots\dots(5)$$

2.5 Koefisien Silhouette

Koefisien *Silhouette* atau *Silhouette Index* merupakan metode untuk mengukur validasi baik sebuah data, *cluster* tunggal atau keseluruhan *cluster*. Nilai *Koefisien Silhouette* data ke-*i* berdasarkan dua komponen yaitu *a_i* dan *b_i*. Komponen *a_i* adalah jarak dari data ke-*i* terhadap semua data dalam *cluster* yang sama. Sedangkan komponen *b_i* adalah rata-rata jarak data ke-*i* terhadap semua data dari *cluster* lain (Prasetyo, 2012).

Berikut adalah rumus untuk menghitung komponen *a_i* :

$$a_i^j = \frac{1}{m_j - 1} \sum_{r \neq i}^{m_j} d(x_i^j, x_r^j), i = 1, 2, \dots, m_j \dots\dots\dots(6)$$

Dimana :

j = *cluster*

i = *index data*

a_i^j = rata-rata jarak ke-*i* terhadap semua data lainnya dalam satu *cluster*

m_j = jumlah data dalam *cluster* ke-*j*

d(x_i^j, x_r^j) = jarak data ke-*i* dengan data ke-*r* dalam satu *cluster* *j*

Berikut rumus menghitung komponen *b_i* :

$$b_i^j = \min_{n \neq j} \left\{ \frac{1}{m_n} \sum_{r \neq i}^{m_n} d(x_i^j, x_r^n) \right\}, i = 1, 2, \dots, m_j \dots\dots\dots(7)$$

Dimana :

j = cluster

n = cluster

i = index data

m_n = banyak data dalam satu cluster

b_i^j = nilai terkecil dari rata-rata jarak data ke- i terhadap semua data dari cluster yang lain tidak dalam satu cluster dengan data ke- i

Untuk menghitung Koefisien Silhouette data ke- i menggunakan persamaan berikut ini:

$$S_i^j = \frac{b_i^j - a_i^j}{\max\{b_i^j, a_i^j\}} \dots\dots\dots(8)$$

Dimana:

S_i^j = Koefisien Silhouette data ke- i dalam satu cluster

a_i^j = rata-rata jarak ke- i terhadap semua data lain dalam satu cluster

b_i^j = nilai terkecil dari rata-rata jarak data ke- i terhadap semua data dari cluster yang lain tidak dalam satu cluster dengan data ke- i

III. METODE PENELITIAN

3.1 Analisa Data

Adapun metode yang digunakan untuk mengolah data adalah sebagai berikut:

3.1.1 Data Selection

Data yang digunakan dalam penelitian adalah data transaksi penjualan hardware periode Januari 2017–Juli 2017 di Delta Computer. Dari semua atribut yang ada di tabel transaksi diambil 3 atribut yang digunakan untuk proses clustering. Atribut tersebut yaitu:

1. Tanggal, merupakan tanggal transaksi penjualan terjadi.
2. Customer, merupakan nama customer yang melakukan transaksi penjualan
3. Nilai, merupakan jumlah biaya pada transaksi penjualan yang dilakukan oleh setiap customer.

3.1.2. Data Pre-Processing

Pada tahapan pre-processing ini akan dilakukan beberapa langkah untuk menghasilkan dataset yang bersih sehingga dapat digunakan dalam tahap berikutnya. Dalam tahapan ini dilakukan penggabungan data dan pembersihan data. Penggabungan data dilakukan untuk menggabungkan data dari transaksi penjualan setiap bulannya menjadi satu tabel. Pembersihan data dilakukan untuk menyortir data yang dapat digunakan untuk proses cluster. Sehingga didapatkan hasil seperti pada Tabel 2.

Tabel 2. Hasil Pre-Processing

Kode	Nama	Recency	Frecuency	Monetary
P001	Adnan Effendi	14/06/2017	1	4.050.000
P002	Andi Lesmana PKU	11/07/2017	1	3.275.000
P003	Anggun	09/06/2017	1	650.000
P004	Angkasa Pura	31/07/2017	1	2.050.000
P005	Anita	03/02/2017	1	150.000
P006	Arif Moewardi	13/05/2017	4	10.465.000
P007	Arsip Sukoharjo	22/07/2017	4	22.965.909
P008	ASMI	28/07/2017	1	16.374.400
P009	Bag. Organisasi	27/07/2017	2	8.150.000
P010	Bag. Perlengkapan UNS	17/06/2017	3	16.370.000

3.1.3. Data Transformation

Pada tahapan ini, data yang telah di seleksi dan diproses akan dilakukan pembobotan *variable* yaitu *Recency*, *Frecuency* dan *Monetary*. Data yang sebelumnya yang masih beragam jenisnya dijadikan data yang sejenis dan siap digunakan untuk proses *cluster*. Aturan pembobotan *variable recency*, *frecuency* dan *monetary* terdapat pada Tabel 3.

Tabel 3. Aturan Pembobotan Variabel

Bobot	R	F	M
4	< 1 bulan	> 10 kali	> 100.000.000
3	2-3 bulan	7-9 kali	50.000.000 – 100.000.000
2	4-5 bulan	4-6 kali	5.000.000 – 50.000.000
1	> 5 bulan	< 4 kali	< 5.000.000

3.2. Proses Clustering

Tahapan yang digunakan untuk perhitungan AHC menggunakan *Average Linkage* adalah sebagai berikut:

1. Hitung matrik kedekatan dengan menggunakan perhitungan jarak *Manhattan Distance* menggunakan rumus nomor 4.
2. Setelah terbentuk matrik diagonal, dicari jarak terdekat antar kelompok data.
3. Gabungkan dua kelompok data terdekat menggunakan metode tautan *Average Linkage* menggunakan rumus nomor 3.
4. Perbarui matriks jarak antar data untuk merepresentasikan kedekatan diantara kelompok baru dan kelompok yang masih tersisa.
5. Ulangi langkah 2 sampai 4 hingga hanya satu kelompok yang tersisa.

3.3. Penerapan Metode Agglomerative Hierarchical Clustering

Berikut adalah penerapan metode AHC yang digunakan untuk perhitungan 10 sampel data.

3.3.1. Perhitungan Matrik Kedekatan dengan Manhattan Distance

Data sampel hasil pembobotan berdasarkan aturan pada Tabel 3, ditunjukkan pada Tabel 4.

Tabel 4. Sampel data

Kode	Recency	Frecuency	Monetary
P001	3	1	1
P002	4	1	1
P003	3	1	1
P004	4	1	1
P005	1	1	1
P006	3	2	2
P007	4	2	2
P008	4	1	2
P009	4	1	2
P010	3	1	2

- Perhitungan jarak P001 dengan P001
 $D_{man} = |3-3| + |1-1| + |1-1| = 0$
- Perhitungan jarak P001 dengan P002
 $D_{man} = |3-4| + |1-1| + |1-1| = 1$
- Perhitungan jarak P001 dengan P003
 $D_{man} = |3-3| + |1-1| + |1-1| = 0$
- Perhitungan jarak P001 dengan P004
 $D_{man} = |3-4| + |1-1| + |1-1| = 1$

- Perhitungan jarak P001 dengan P005
 $D_{man} = |3-1| + |1-1| + |1-1| = 2$
- Perhitungan jarak P001 dengan P006
 $D_{man} = |3-3| + |1-2| + |1-2| = 2$
- Perhitungan jarak P001 dengan P007
 $D_{man} = |3-4| + |1-2| + |1-2| = 3$
- Perhitungan jarak P001 dengan P008
 $D_{man} = |3-4| + |1-1| + |1-2| = 2$
- Perhitungan jarak P001 dengan P009
 $D_{man} = |3-4| + |1-1| + |1-2| = 2$
- Perhitungan jarak P001 dengan P010
 $D_{man} = |3-3| + |1-1| + |1-2| = 1$

Perhitungan jarak dilakukan seterusnya dari P002 dengan P001 hingga P010 dan selanjutnya sama untuk jarak antar data. Setelah seluruh data dari P001 hingga P010 sudah dihitung jarak antar data akan terbentuk matrik diagonal seperti pada Tabel 5.

Tabel 5. Matrik Diagonal

D_{man}	1	2	3	4	5	6	7	8	9	10
1	0	1	0	1	2	2	3	2	2	1
2	1	0	1	0	3	3	2	1	1	2
3	0	1	0	1	2	2	3	2	2	1
4	1	0	1	0	3	3	2	1	1	2
5	2	3	2	3	0	4	5	4	4	3
6	2	3	2	3	4	0	1	2	2	1
7	3	2	3	2	5	1	0	1	1	2
8	2	1	2	1	4	2	1	0	0	1
9	2	1	2	1	4	2	1	0	0	1
10	1	2	1	2	3	1	2	1	1	0

3.3.2. Perhitungan Average Linkage

Setelah terbentuk matrik diagonal, dicari pasangan data yang memiliki jarak antar data yang terkecil menggunakan rumus (3) yaitu pasangan data 1 dan 3.

$$\text{Min}(D_{man}) = \min d_{(13)} = 0$$

Pasangan data 1 dan 3 memiliki jarak terkecil, sehingga data 1 dan 3 digabungkan.

Selanjutnya menghitung jarak antara pasangan data (1 dan 3) dengan data yang lainnya 2, 4, 5, 6, 7, 8, 9 dan 10 menggunakan rumus 6.

$$d_{(13)2} = \text{average}(d_{12}, d_{32}) = \frac{1+1}{2} = 1$$

$$d_{(13)2} = \text{average}(d_{12}, d_{32}) = \frac{1+1}{2} = 1$$

$$d_{(13)4} = \text{average}(d_{14}, d_{34}) = \frac{1+1}{2} = 1$$

$$d_{(13)5} = \text{average}(d_{15}, d_{35}) = \frac{2+2}{2} = 2$$

$$d_{(13)6} = \text{average}(d_{16}, d_{36}) = \frac{2+2}{2} = 2$$

$$d_{(13)7} = \text{average}(d_{17}, d_{37}) = \frac{3+3}{2} = 3$$

$$d_{(13)8} = \text{average}(d_{18}, d_{38}) = \frac{2+2}{2} = 2$$

$$d_{(13)9} = \text{average}(d_{19}, d_{39}) = \frac{2+2}{2} = 2$$

$$d_{(13)10} = \text{average}(d_{110}, d_{310}) = \frac{1+1}{2} = 1$$

Langkah berikutnya adalah menghapus baris dan kolom matrik yang bersesuaian dengan data 1 dan 3, serta menambahkan baris dan kolom untuk pasangan data (13).

Tabel 6. Tabel Matrik Jarak Pertama

D_{man}	(13)	2	4	5	6	7	8	9	10
(13)	0	1	1	2	2	3	2	2	1
2	1	0	0	3	3	2	1	1	2
4	1	0	0	3	3	2	1	1	2
5	2	3	3	0	4	5	4	4	3
6	2	3	3	4	0	1	2	2	1
7	3	2	2	5	1	0	1	1	2
8	2	1	1	4	2	1	0	0	1
9	2	1	1	4	2	1	0	0	1
10	1	2	2	3	1	2	1	1	0

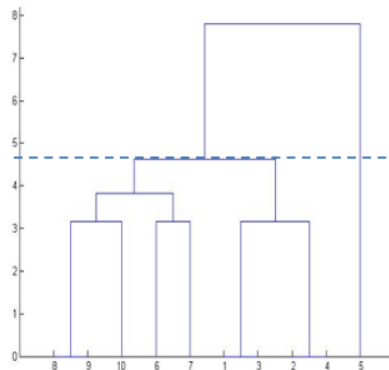
Ulangi langkah tersebut sampai ditemukan pasangan data atau kelompok data seperti Tabel 7.

Tabel 7. Tabel Matrik Jarak Kedelapan

D_{man}	(1324891067)	5
(1324891067)	0	3.3
5	3.3	0

4

Jadi pasangan data (1324891067) dan 5 digabung untuk menjadi kelompok tunggal dari 10 data, yaitu kelompok (12345678910) dengan jarak terdekat 3.3. Berikut ini adalah hasil dari metode *Average Linkage* dengan perpotongan (*cut off*) di titik 4.7, sehingga *cluster* yang terbentuk ada tiga, dengan rincian seperti pada Tabel 8.



Gambar 1. Dendrogram

Tabel 8. Hasil Clustering

Cluster	Jumlah	Anggota Cluster	Avg R	Avg F	Avg M
1	5 anggota	6,7,8,9,10	3.5	1	1
2	4 anggota	1,3,2,4	3.6	1.4	2
3	1 anggota	5	1	1	1

3.4. Pengujian Koefisien *Silhouette*

Setelah didapatkan hasil 3 *cluster* akan dilakukan pengujian validitas pada setiap data untuk mengetahui apakah hasil *cluster* merupakan hasil yang terbaik atau masih belum maksimal. Pengujian validitas menggunakan Koefisien *Silhouette* dengan rumus 7, 8 dan 9.

- Data ke-1

$$a_1^2 = \frac{1}{4-1} (0 + 1 + 0 + 1) = \frac{1}{3} (2) = 0,6667$$

$$b_1^2 = \min\left(\frac{1}{1}, 2\right) \left(\frac{1}{5} \cdot (1,4142 + 1,7321 + 1,4142 + 1,4142 + 1)\right)$$

$$= \min(2)(1,39494) = 1,39494$$

$$s_1^2 = \frac{1,39494 - 0,6667}{1,39494} = \frac{0,72824}{1,39494} = 0,5221$$

• Data ke-2

$$a_2^2 = \frac{1}{4-1} (1 + 0 + 1 + 0) = \frac{1}{3} (2) = 0,6667$$

$$b_2^2 = \min\left(\frac{1}{1}, 2\right) \left(\frac{1}{5} \cdot (1,7321 + 1,4142 + 1 + 1 + 1,4142)\right)$$

$$= \min(2)(1,31209) = 1,31209$$

$$s_2^2 = \frac{1,31209 - 0,6667}{1,31209} = \frac{0,64539}{1,31209} = 0,4919$$

Untuk data ke-3 sampai data ke-10 dilakukan perhitungan yang sama seperti diatas sehingga diperoleh hasil seperti pada Tabel 9. Menunjukkan bahwa *cluster 1*, *cluster 2* dan *cluster 3* menunjukkan kualitas hasil *cluster* yang cukup baik karena tidak memiliki nilai minus.

Tabel 9. Hasil Pengujian Koefisien *Silhouette*

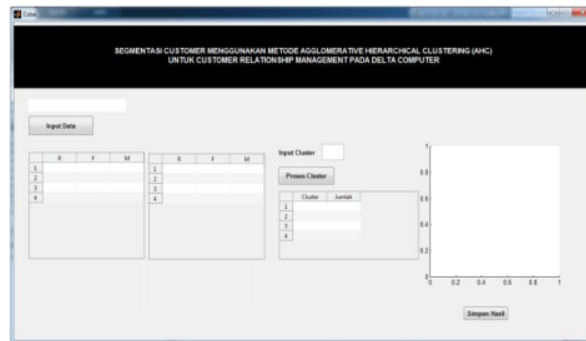
Data Ke-	Koef. Silhouette
1	0,5221
2	0,4919
3	0,5221
4	0,4919
5	1
6	0,2327
7	0,2985
8	0,2929
9	0,2929
10	0,0858

11

IV. HASIL DAN PEMBAHASAN

4.1 Implementasi *User Interface*

Implementasi sistem Segmentasi *Customer* Menggunakan *Agglomerative Hierrarchical Clustering* untuk *Customer Relationship Management* dapat dilihat pada Gambar 2. Dari hasil *clustering* tersebut dapat diidentifikasi karakteristik *customer* berdasarkan tingkat kelayolitasannya dengan perhitungan nilai RFM (Hardiani et al., 2015).



Gambar 2. Implementasi *User Interface*

Dari hasil implementasi *user interface* diatas berikut adalah penjelasan detailnya:

a. *Input data*

Menu *input data* digunakan untuk memasukkan data penjualan yang berekstensi .xls atau .xlsx yang akan diolah untuk yang terdiri dari 3 *variable* dan setelah *diinput data* otomatis akan di bobotkan sesuai aturan pembobotan.

b. Proses *Clustering*

Setelah data *diinputkan* dan dibobotkan, selanjutnya adalah melakukan proses *clustering* dengan menginputkan jumlah cluster yang diinginkan lalu menekan tombol proses *cluster*, lalu hasil dari proses *clustering* akan tampil pada tabel dan hasil dendrogram akan tampil pada axes.

c. Simpan Hasil

Setelah proses *clustering* selesai data dapat disimpan dengan menekan tombol simpan hasil, lalu memilih lokasi untuk menyimpan *file* dan memberi nama *file*. *File* yang tersimpan berupa *file Excel* yang berisikan hasil proses *clustering* dan hasil data yang disimpan digunakan untuk proses selanjutnya yaitu penentuan jenis *customer* dan penentuan strategi promosi untuk setiap *cluster*.

4.2 Analisa Hasil

Hasil pengujian Koefisien *Silhouete* apabila hasil semakin mendekati 1 maka hasil akan semakin baik, sebaliknya jika hasil mendekati 0 atau bahkan minus menunjukkan bahwa hasil dari penempatan *cluster* dari data ke-*i* kurang tepat. Setelah dihitung koefisien *Silhouette* dari seluruh data *customer* yang berjumlah 308 data pada 4 *cluster*, 6 *cluster* dan 7 *cluster*, dapat disimpulkan bahwa 7 *cluster* adalah hasil terbaik dalam proses *clustering* tersebut. Hal ini didasarkan pada hasil minus yang paling sedikit di 7 *cluster* dengan 6 anggota, yang ditunjukkan pada Tabel 10.

Tabel 10. Hasil perbandingan validitas *cluster*

Hasil Cluster	10 4 Cluster	6 Cluster	7 5 Cluster
Jumlah anggota cluster	Cluster 1 = 9 anggota Cluster 2 = 2 anggota Cluster 3 = 1 an 10 ta Cluster 4 = 296 anggota	Cluster 1 = 75 anggota Cluster 2 = 158 anggota Cluster 3 = 63 anggota Cluster 4 = 9 anggota Cluster 5 = 2 anggota Cluster 6 = 1 anggota	Cluster 1 = 20 anggota Cluster 2 = 43 anggota Cluster 3 = 75 anggota Cluster 4 = 158 anggota Cluster 5 = 9 anggota Cluster 6 = 2 anggota Cluster 7 = 1 anggota
Jumlah hasil minus pada cluster	17 hasil minus : - 2 dari cluster 1 - 15 dari cluster 4	8 hasil mi 17 : - 1 dari cluster 1 - 1 dari cluster 2 - 3 dari cluster 3 - 3 dari cluster 4	6 hasil minus : - 1 dari cluster 3 - 1 dari cluster 4 - 4 dari cluster 5

Dari 7 *cluster* yang terbentuk berikut pada Tabel 11, dapat dilihat jumlah anggota dan rata-rata *Recency*, *Frecuency* dan *Monetary* dari setiap *cluster*.

Tabel 11. Hasil Proses *Clustering*

Cluster	Jumlah Anggota	Avg R	Avg F	Avg M
1	20	4	1,1	1
2	43	3,6976	1,3720	2,2325
3	75	1	1	1,4666
4	158	2,4493	1,0253	1,5063
5	9	3,6666	2,6666	3
6	2	1,5	1,5	4
7	1	4	4	4

Setelah diketahui karakteristik *customer* dari *Recency*, *Frecuency* dan *Monetary*, selanjutnya dilakukan penentuan strategi promosi yang tepat sesuai dengan jenis *customer*. Penentuan strategi promosi yang tepat berdasarkan jumlah *cluster* dan karakteristik *customer* dapat dilihat pada Tabel 12.

Tabel 12. Strategi Promosi

Cluster	Jenis Customer	Strategi Promosi
Cluster 1	<i>Everyday Shopper</i>	Memberikan promo-promo menarik secara berkala.
Cluster 2	<i>Typical Customer</i>	Pemberian potongan harga atau diskon.
Cluster 3	<i>Dormant Customer</i>	Meningkatkan pola pemasaran melalui website, media sosial dan brosur.
Cluster 4	<i>Typical Customer</i>	Pemberian potongan harga atau diskon.
Cluster 5	<i>Golden Customer</i>	Pemberian <i>reward</i> .
Cluster 6	<i>Occasional Customer</i>	Pemberian bonus atau barang.
Cluster 7	<i>Superstar</i>	Pemberian kartu <i>member</i> .

11
V.
5.1

PENUTUP
Kesimpulan

Berdasarkan penelitian ini, maka dapat diambil beberapa kesimpulan yaitu:

1. *Agglomerative Hierarchical Clustering* (AHC) dapat digunakan untuk pengelompokkan pelanggan dan dikaitkan dengan strategi pemasaran.
2. Hasil pengelompokkan diperoleh *cluster 1* terdapat 20 anggota, *cluster 2* terdapat 43 anggota, *cluster 3* terdapat 75 anggota, *cluster 4* terdapat 158 anggota, *cluster 5* terdapat 9 anggota, *cluster 6* terdapat 2 anggota dan *cluster 7* terdapat 1 anggota.
3. Pengujian hasil menggunakan metode *Koefisien Silhouette* mendapatkan hasil dalam *cluster 1* terdapat 20 data tepat berada di *cluster 1* dari 20 data, dalam *cluster 2* terdapat 43 data yang tepat berada di *cluster 2* dari 43 data, dalam *cluster 3* terdapat 74 data yang tepat berada di *cluster 3* dari 75 data, dalam *cluster 4* terdapat 157 data yang tepat berada di *cluster 4* dari 158 data, dalam *cluster 5* terdapat 5 data yang tepat berada di *cluster 5* dari 9 data, dalam *cluster 6* terdapat 2 data yang tepat berada di *cluster 6* dari 2 data dan dalam *cluster 7* hanya terdapat 1 data.
4. Karakteristik *customer* yang terbentuk dalam setiap *cluster* yaitu *cluster 1* termasuk *Everyday Shopper*, *cluster 2* dan *cluster 4* termasuk *Typical Customer*, *cluster 3* termasuk *Dormant Customer*, *cluster 5* termasuk *Golden Customer*, *cluster 6* termasuk *Occasional Customer* dan *cluster 7* termasuk *Superstar*.
5. Dalam menjaga hubungan yang baik dengan pelanggan strategi promosi yang diterapkan pada setiap karakteristik pelanggan yaitu untuk *Everyday Shopper* memberikan promo-promo menarik secara berkala, *Typical Customer* pemberian potongan harga atau diskon, untuk *Dormant Customer* peningkatan pemasaran melalui web, sosial media dan brosur, untuk *Superstar* pemberian kartu *member*, untuk *Golden Customer* pemberian *reward* serta untuk *Occasional Customer* pemberian bonus atau barang.

5.2 **Saran**

Penelitian berikutnya yang dapat dilakukan adalah sebagai berikut:

1. Perhitungan metode *Agglomerative Hierarchical Clustering* dapat dilakukan dengan menggunakan metode selain *Average Linkage* seperti *Complete Linkage*, *Single Linkage* maupun *Ward* serta pengukuran jarak dapat menggunakan *Euclidean* ataupun yang lain.

2. Pengujian validasi dapat dicoba menggunakan perhitungan selain Koefisien *Silhouette*.

DAFTAR PUSTAKA

- Effendi, A. L. (2015). Segmentasi Konsumen Sayuran Organik Pada Yogya Dept. Store Kota Bandung Tahun 2014. *Jurnal Telekomunikasi, Bisnis Informatika, Dan Manajemen*, 2(1), 10–17.
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5th Editio). The Atrium, Southern Gate, Cichester, West Sussex, P019 8SQ: John Wiley & Son.
- Greenberg, J., & Baron, J. A. (2010). *Behavior in Organizations: Understanding and Managing The Human Side of work*. New Jersey: Pearson Education International.
- Hardiani, T., Sulisty, S., & Hartanto, R. (2015). Segmentasi Nasabah Tabungan Menggunakan Model RFM (Recency, Frecuency, Monetary) dan K-Means pada Lembaga Keuangan Mikro. In *Seminar Nasional Teknologi Informasi dan Komunikasi Terapan (SEMANTIK)*.
- Kurniawan, A. A. (2017). *Implementasi Algoritma Agglomerative Hierarchical Clustering Untuk Mengelompokkan Capaian Belajar Siswa SD*. Universitas Sanata Dharma.
- Laeli, S. (2014). *Analisis Cluster dengan Average Linkage Method dan Ward's Method untuk Data Responden Nasabah Asuransi Jiwa Unit Link*. Universitas Negeri Yogyakarta.
- Lubis, A. H. (2016). Model Segmentasi Pelanggan Dengan Kernel K-Means Clustering Berbasis Customer Relationship Management. *Sinkron*, 1(1).
- Prasetyo, E. (2012). *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi.
- Rainer, R. K., & Cegielski, C. G. (2011). *Introduction to Information Systems:Enabling and Transforming Business* (3rd Editio). USA: John Wiley & Sons.
- Sulistiyowati, Ketherin, B. E., Arifiyanti, A. A., & Sodik, A. (2018). Analisa Segmentasi Konsumen Menggunakan Algoritma K-Means Clustering. In *Seminar Nasional Sains dan Teknologi Terapan: Pendekatan Multidisiplin Menuju Teknologi dan Industri yang Berkelanjutan* (pp. 51–58). Surabaya: Institut Teknologi Adhi Tama Surabaya.
- Vulandari, R. T. (2017). *Data Mining*. Yogyakarta: Gava Media.
- Vulandari, G. F. (2014). *Segmentasi Pelanggan Menggunakan Algoritma K-Means Untuk Customer Relationship Management (CRM) pada Miulan Hijab*. Universitas Dian Nuswantoro.

Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan

ORIGINALITY REPORT

16%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.researchgate.net Internet	199 words — 4%
2	text-id.123dok.com Internet	110 words — 2%
3	repository.usd.ac.id Internet	84 words — 2%
4	eprints.binadarma.ac.id Internet	79 words — 2%
5	www.unisbank.ac.id Internet	47 words — 1%
6	repository.uin-suska.ac.id Internet	40 words — 1%
7	repository.usu.ac.id Internet	38 words — 1%
8	download.garuda.kemdikbud.go.id Internet	35 words — 1%
9	core.ac.uk Internet	27 words — 1%

10	ejournal.nusamandiri.ac.id Internet	23 words — < 1%
11	docplayer.info Internet	22 words — < 1%
12	Dwi Remawati, Paulus Harsadi, Ruyyanto Dwi Nugroho. "Penerapan Sistem Penunjang Keputusan Menggunakan Algoritma Naive Bayes Pada konsep Human Resource Information System (HRIS) (Studi kasus :Penerusan Kontrak Kerja Karyawan di PT. XYZ)", Jurnal Ilmiah SINUS, 2020 Crossref	17 words — < 1%
13	id.123dok.com Internet	15 words — < 1%
14	adoc.pub Internet	14 words — < 1%
15	garuda.kemdikbud.go.id Internet	13 words — < 1%
16	jurnalsaintek.uinsby.ac.id Internet	13 words — < 1%
17	anzdoc.com Internet	10 words — < 1%
18	eprints.umm.ac.id Internet	10 words — < 1%
19	media.neliti.com Internet	10 words — < 1%

EXCLUDE QUOTES ON

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES < 3 WORDS

EXCLUDE MATCHES < 10 WORDS