



ANALISA METODE *HIERARCHICAL CLUSTERING* DAN *K-MEAN* DENGAN MODEL LRFMP PADA SEGMENTASI PELANGGAN

Asep Muhidin

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa
asep.muhidin@pelitabangsa.ac.id

Abstrak

Pelanggan adalah sesuatu yang berharga dan penting, jika semua pelanggan serupa, bisnis akan begitu sederhana. Masalah heterogenitas dan banyaknya jumlah pelanggan menjadi tantangan yang harus dihadapi untuk menentukan segmentasi konsumen yang potensial. Pada penelitian ini proses segmentasi pelanggan dimulai dengan melakukan proses preprocessing, analytic hierarchy process (AHP), pencarian nilai K terbaik dari semua metode Hierarchical Clustering dengan membandingkan nilai Bouldien-Index. Selanjutnya nilai K terpilih dijadikan nilai awal pada K-Mean Clustering. Hasil clustering tersebut digunakan untuk melakukan segmentasi menggunakan model RFM untuk mendapatkan kelas konsumen. Penambahan parameter Payment (LRFMP) dapat meningkatkan nilai loyalitas pelanggan terhadap perusahaan. Berdasarkan hasil penelitian, metode single linkage merupakan metode terbaik untuk mencari nilai K. Segmentasi model k-mean dengan penambahan parameter P (LRFMP) dapat meningkatkan nilai DBI dibandingkan dengan model RFM terbobot maupun tidak. Tetapi nilai DBI metode segmentasi single linkage masih lebih bagus dari pada segmentasi k-mean.

Kata Kunci: *Bouldien-Index, CRM, Data mining, pelanggan, LRFMP, RFM, segmentasi*

Abstract

Customer is something valuable and important, if all customers are similar, business will be so simple. The problem of heterogeneity and the large number of customers is a challenge to be faced in determining potential customer segmentation. In this research, customer segmentation process begins with preprocessing process, analytic hierarchy process (AHP), search for the best value of all Hierarchical Clustering methods by comparing the Bouldien-Index value. Furthermore, the value of K is chosen to be the initial value in K-Mean Clustering. The clustering result is used to segment the RFM model to get the consumer class. The addition of Payment parameter (LRFMP) can increase the value of customer loyalty to the company. Based on the research results, the single linkage method is the best method to find the value of K. Segmentation of the k-mean model with the addition of the P (LRFMP) parameter can increase the DBI value compared to the weighted RFM model or not. But the DBI value of the single linkage segmentation method is still better than the k-mean segmentation.

Keyword: *bouldien-Index, clustering, CRM, customer, data mining, hierarchical clustering, LRFMP, preprocessing, RFM, segmentation.*

1. Pendahuluan

Pemahaman terhadap pelanggan adalah sesuatu yang berharga dan penting. Jika semua pelanggan serupa, bisnis akan begitu sederhana. Analisa segmentasi yang baik terhadap pelanggan dapat digunakan untuk memetakan pelanggan potensial dan tidak potensial. Data transaksi pelanggan berkembang sangat pesat, sehingga volume data semakin hari semakin besar dan kompleks, baik dari jumlah *record* maupun jumlah *field*. Dengan perkembangan data tersebut tidak mungkin menganalisa perilaku pelanggan dengan metode tradisional atau manual. Penggunaan metode-metode data mining dapat digunakan untuk menganalisa data yang diperoleh dari transaksi, sehingga dapat menggali informasi-informasi tersembunyi dari data transaksi tersebut, dan salah satunya adalah informasi segmentasi pelanggan.

Data mining adalah proses menemukan pengetahuan, ketertarikan, pola yang tidak diketahui yang dapat dijelaskan, dimengerti dan diprediksi dari data skala besar[1]. Data mining adalah proses otomatis menemukan informasi yang berguna dalam repositori data yang besar. Salah satu teknik data mining yang digunakan untuk mencari segmentasi pelanggan adalah teknik clustering. *Clustering* adalah proses pengelompokan sekumpulan obyek data ke dalam beberapa kelompok atau *cluster* sehingga obyek dalam sebuah *cluster* memiliki kemiripan yang tinggi, tetapi sangat berbeda dengan obyek dalam kelompok lainnya[2]. Ada dua metode *clustering*, yaitu *Hierarchical Clustering* dan *Non Hierarchical Clustering*. Metode *Hierarchical Clustering* terdiri dari *Single Linkage Clustering*, *Complete Linkage Clustering*, *Average Linkage Clustering*, *Centroid Linkage Clustering* dan metode *Ward*. Sedangkan metode *Non Hierarchical Clustering* sendiri terdiri dari *K-means* dan *Fuzzy K-means*.

Penelitian ini bertujuan menganalisis metode *Hierarchical Clustering* dan *K-Means* pada pengelompokan karakteristik pelanggan berdasarkan model LRFMP (*Length, Recency, Frequency, Monetary, Payment*).

2. Kajian Pustaka

Penelitian segmentasi pelanggan dengan metode clustering dan model RFM pernah dilakukan sebelumnya. Parvaneh, M. J. Tarokh dan H. Abbasimehr (2014) melakukan penelitian dengan algoritma K-Mean. Preprocessing data yaitu *AHP* (*Analytic Hierarchy Process*), pembobotan nilai LRFMP berdasarkan nilai *AHP* dan menentukan nilai *K* optimum menggunakan metode *Davies Index-Bouldin*. Penelitian tersebut untuk mencari segmen pelanggan potensial, dengan menambahkan parameter *P* (*Potensial*) pada model RFM.

Mohsen Alvandi, Safar Fazli dan Farzaneh Seifi Abdoli (2012) melakukan penelitian dengan metodologi yang digunakan yaitu *CRSIP* (*Cross Industry Standard Process for Data Mining*). Metodologi *CRSIP* meliputi : *Business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation* dan *deployment*. Metode pengelompokan menggunakan *K-mean* dan metode validasi *RS* (*R Squared*) *Index*.

2.1. Metode Hirarki

Metode hirarki (*hierarchical methods*) adalah salah satu teknik clustering dengan cara mengelompokkan dua atau lebih obyek yang mempunyai nilai kemiripan yang paling dekat. Selanjutnya hasil pengelompokan pertama dikelompokkan lagi dengan obyek lain yang mempunyai nilai kemiripan kedua. Demikian seterusnya, sehingga akan membentuk konstruksi hirarki atau berdasarkan tingkatan tertentu seperti struktur pohon (struktur pertandingan). Dendogram

digunakan untuk menggambarkan struktur hasil cluster hirarki tersebut.

Beberapa Metode-metode Teknik cluster hirarki yang digunakan adalah sebagai berikut :

- *Single Linkage (Nearest Neighbor Methods)* : metode ini mendefinisikan jarak antara dua kelompok menjadi jarak minimum antara setiap titik data tunggal di cluster pertama dan setiap titik data tunggal di *cluster* kedua.
- *Complete Linkage (Furthest Neighbor Methods)* : metode ini mencari nilai jarak maximum antar obyek. Dimulai dengan pencarian dua obyek yang paling dekat jaraknya dan kedua obyek tersebut membentuk *cluster* yang baru (pertama).
- *Average Linkage Methods (Between Groups Methods)* : metode ini mencari nilai jarak rata-rata antar obyek. Dimulai dengan pencarian dua obyek yang paling dekat jaraknya dan kedua obyek tersebut membentuk cluster yang baru (pertama).
- *Centroid Method* : Pada metode centroid method, jarak antara dua cluster terbentuk adalah jarak diantara dua centroid pada kedua cluster tersebut. *Centroid* adalah rata-rata jarak antar obyek pada sebuah cluster. Nilai *centroid* didapat dengan melakukan rata-rata pada semua anggota cluster. Pada metode ini, proses perhitungan ulang centroid akan dilakukan setiap terbentuk *cluster* baru, sampai terbentuk cluster yang tetap.

2.2. Metode K-Mean

K-Means adalah salah satu metode clustering non hirarki yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode ini melakukan analisis kelompok yang mengarah pada pembagian *M* obyek pengamatan ke dalam *C* kelompok (*cluster*). Metode ini mencoba untuk menemukan pusat dari kelompok (*centroid*) dalam data sebanyak iterasi perbaikan yang dilakukan. Metode ini berusaha membagi data kedalam kelompok sehingga data yang berkarakteristik sama dimasukkan ke dalam satu kelompok, sementara data yang berkarakteristik berbeda dimasukkan dalam kelompok yang lain. Secara umum Algoritma Dasar dari *K-Means Clustering* adalah sebagai berikut :

- 1) Menentukan jumlah *cluster*
- 2) Menempatkan data ke dalam kelompok secara acak
- 3) Menghitung pusat cluster (*centroid*) dengan mencari nilai rata-rata (*mean*) untuk masing-masing kelompok
- 4) Tempatkan data ke centroid terdekat
- 5) Kembali ke langkah 3, jika masih ada data yang berpindah cluster atau jika nilai centroid diatas nilai ambang, atau jika nilai pada fungsi obyektif yang digunakan masih diatas ambang.

2.3. Indeks Davies-Bouldin (IDB)

Metode ini bertujuan memaksimalkan pengukuran jarak cluster antara satu cluster dengan cluster yang lain, juga meminimalkan jarak antar anggota dalam satu cluster. Sebelum menghitung *Davies-Bouldin Index (DBI)*, dihitung terlebih dahulu variance dari masing-masing cluster.

$$Var(x) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

dimana :

- \bar{x} : rata-rata dari cluster x dan
- N : jumlah anggota cluster.

Kemudian menghitung *Davies-Bouldin Index (DBI)* dengan persamaan

$$DBI = \frac{1}{k} \sum_{i=1}^k R_i$$

$$R_i = \max_{j=1..k, i \neq j} R_{ij}$$

$$R_{ij} = \frac{\text{var}(C_i) + \text{var}(C_j)}{\|c_i - c_j\|}$$

Dimana :

C_i = cluster i dan c_i adalah centroid dari cluster I

2.4. Analytic Hierarchy Process (AHP)

Prof. Thomas L.Saaty mengembangkan suatu metode algoritma pengambil keputusan untuk menyelesaikan permasalahan multikriteria. Metode tersebut dikenal dengan metode *Analytic Hierarchy Process* atau *AHP*. *AHP* menyederhanakan permasalahan multikriteria kedalam bentuk hierarki yang terdiri dari 3 komponen utama yaitu tujuan atau goal dari pengambilan keputusan, kriteria penilaian dan alternatif pilihan. Tabel penetapan prioritas elemen dengan perbandingan berpasangan digunakan untuk skala perbandingan sebagai ukuran yang menyatakan intensitas kepentingan (Saaty 1988).

Tabel 1. Penetapan prioritas elemen dengan perbandingan berpasangan

Intensitas Kepentingan	Keterangan	Penjelasan
1	Kedua elemen sama pentingnya	Dua elemen mempunyai pengaruh yang sama besar terhadap tujuan
3	Elemen yang satu sedikit lebih penting dari pada elemen yang lainnya	Pengalaman dan penilaian sedikit menyokong satu elemen dibandingkan elemen lainnya
5	Elemen yang satu lebih penting dari pada elemen yang lainnya	Pengalaman dan penilaian sangat kuat menyokong satu elemen dibandingkan elemen lainnya
7	Satu elemen jelas lebih penting dari pada elemen lainnya	Satu elemen yang kuat dikosong san dominan terlihat dalam praktek
9	Satu elemen mutlak penting dari pada elemen lainnya	Bukti yang mendukung elemen yang satu terhadap elemen lain memiliki tingkat penegasan tertinggi yang mungkin menguatkan
2,4,6,8	Nilai-nilai antara dua nilai pertimbangan yang berdekatan	Nilai ini diberikan bila ada dua kompromi diantara dua pilihan
Kebalikan	Jika untuk aktivitas i mendapat satu angka dibanding dengan aktivitas j, maka j mempunyai nilai kebalikannya dibanding dengan i.	

3. Metode Penelitian

Analisa segmentasi pelanggan pada penelitian ini mengunakan tahapan yang sistematis dan terarah ,sehingga dapat diketahui metode manakah yang menghasilkan hasil *cluster* yang terbaik. Tahapan yang dilakukan adalah : pemahaman dan *preprocessing* data , *RFM modeling*, penggunaan *Analytical Hierarchy Process (AHP)* untuk pembobotan model *LRFMP*, menentukan nilai K optimal berdasarkan *Davies-Bouldin Index*, klasterisasi dengan *K-Mean Clustering* dan Uji F Anova (*Analysis of Variance*) untuk hasil *Clustering K-Mean*

3.1. Pemahaman dan *Preprocessing* Data

Database perusahaan meliputi berbagai macam data, diantara data yang ada terdapat data pelanggan dan transaksi pelanggan yang memenuhi kebutuhan penelitian untuk segmentasi pelanggan.Data transaksi dan data pelanggan yang didapatkan melalui *pre-processing* untuk mendapatkan data yang benar-benar sesuai dengan keperluan penelitian. *Pre-processing* yang dilakukan adalah *Data Selection* : yaitu pemilihan atribut-atribut yang sesuai dengan proses data mining. *Data Preprocessing* : adalah pembersihan data (*cleansing*) dan *missing values* dengan melakukan metode-metode *query* sederhana untuk menemukan anomali- anomali data yang bisa saja masih terdapat pada sistem. *Data Transformation*: mengelompokkan atribut-atribut yang telah terpilih menjadi satu tabel dengan cara melakukan denormalisasi dari basis data.Standarisasi Data : standarisasi variabel dilakukan apabila terdapat perbedaan satuan yang signifikan diantara variabel-variabel yang diteliti.

3.3. RFM Modeling

Model RFM adalah model yang biasa digunakan untuk menilai loyalitas pelanggan, model ini dikenalkan oleh Highes(1994), yang dibangun berdasarkan tiga kriteria yaitu :

- Recency* : mengacu pada durasi waktu/periode pembelian terakhir dengan saat ini; nilai yang lebih rendah sesuai dengan probabilitas yang lebih tinggi dari pelanggan yang melakukan pembelian secara berulang.
- Frequency* : banyaknya jumlah pembelian yang dilakukan dalam jangka waktu/ periode tertentu; frekuensi yang lebih tinggi menunjukkan loyalitas yang lebih besar.
- Monetary* : nilai total pembelian yang dihabiskan selama periode tertentu; jumlah yang lebih tinggi menunjukkan kontribusi yang besar bagi perusahaan.

Pada penelitian ini penulis mengusulkan penambahan variable baru yaitu P (*Payment*). P adalah nilai jumlah *overdue* (pembayaran lewat jatuh tempo) yang dilakukan oleh customer. Pembayaran tepat waktu adalah salah ukuran loyalitas pelanggan terhadap perusahaan.

3.4. Penggunaan Analytical Hierarchy Process (AHP) untuk Pembobotan Model LRFMP

Metode AHP digunakan untuk menentukan bobot prioritas kriteria model LRFMP dinotasikan dengan w_L, w_R, w_F, w_M dan w_P . Metode AHP diimplementasikan pada hasil penilaian pihak marketing dalam menilai pengaruh kriteria length, recency, frequency, monetary, dan payment terhadap operasional perusahaan. Langkah selanjutnya menghitung nilai index inkonsistensi dan diperiksa untuk setiap penilaian tersebut, dan akhirnya didapat nilai bobot tiap kriteria. Bobot tersebut dikalikan dengan nilai kriteria model LRFMP sebagai penentu peringkat klaster dengan Persamaan :

$$\text{LRFMP terbobot} = w_L \times L + w_R \times R + w_F \times F + w_M \times M + w_P \times P$$

Semakin tinggi peringkatnya, maka semakin besar tingkat kelayakan pelanggan dalam klaster tersebut.

3.4. Menentukan nilai K optimal berdasarkan Davies-Bouldin Index

Data hasil AHP disegmentasi dengan seluruh metode *Hierarchical Clustering*, yaitu dari *Complete Linkage, Single Linkage, Average Linkage Clustering, Centroid* dan metode *Wards*. Hasil segmentasi yang terbentuk akan dievaluasi menggunakan *Davies-Bouldin (DB) Index*. Semakin kecil nilai DB Index menunjukkan skema *cluster* yang paling optimal.

3.5. Klasterisasi dengan K-Mean Clustering

Tahap klasterisasi menggunakan algoritma *K-Mean Clustering*. Data yang digunakan sebagai

masukan adalah data nilai LRFM normalisasi dengan bobot dan LRFMP normalisasi dengan bobot. Sedangkan nilai K awal adalah nilai K hasil *analisa Davies Bouldin Index*.

3.5. Uji F Anova (Analysis of Variance) Hasil Clustering K-Mean

Uji ANOVA untuk mengetahui apakah cluster-cluster yang dihasilkan berbeda satu sama lain. Rumus F Anova adalah :

$$F = \frac{SSB}{SSW}$$

$$SSW(C, m) = \frac{1}{n} \sum_{i=1}^m \sum_{j \in C_i} \|x_j - C_{(p)j}\|$$

$$SSB(C, m) = \frac{1}{n} \sum_{i=1}^m n_i \|C_i - \bar{x}\|$$

di mana m banyaknya cluster, C_i adalah *cluster* yang terbentuk, $i=1, \dots, m$, n_i , yaitu banyaknya obyek *cluster* ke- i , x_{ij} merupakan obyek ke- j pada cluster ke- i dan $C_{(p)j}$ merupakan perkalian $x_{ij} n_i$. Semakin besar nilai F , maka semakin besar perbedaan variabel pada cluster yang terbentuk.

4. Hasil dan Pembahasan

Data yang akan dijadikan dataset penelitian merupakan data perusahaan XYZ. Data-data yang diproses menjadi dataset adalah data tabel *Customer, SOHeader, SODetail, Invoice_H, Invoice_D, PayAR_H* dan *PayAR_D*. Operasi dengan perintah SQL terhadap tabel-tabel tersebut menghasilkan kolom-kolom yang diperlukan dalam penelitian, yaitu:

- CustID : Data kode customer
- L : Data perbedaan waktu transaksi pertama dan terakhir dalam satuan tahun
- R : Data perbedaan waktu transaksi terakhir dengan waktu penelitian dalam satuan bulan
- F : Data jumlah (*frekuensi*) *order* customer
- M : Jumlah uang (*amount*) yang dikeluarkan dalam keseluruhan order
- P : Data pembayaran tidak tepat waktu

Jumlah record yang diteliti sebanyak 267 record

Tabel 2. Data transaksi dan pelanggan

CustID	L	R	F	M	P
2596	1	1	1	8	0
2581	1	1	2	29	0
2627	1	1	3	6	0
2648	1	1	3	35	1
2620	1	1	10	167	0
2638	1	1	16	74	2
2597	1	1	53	34527	97
2606	1	2	1	8	0
487	1	2	1	8	0
2617	1	2	10	711	1
2615	1	3	1	11	0
2588	1	3	3	337	0
2616	1	3	4	53	0
2643	1	4	1	10	0
2641	1	4	2	30	1
2631	1	4	2	50	0
2610	1	4	5	234	0
2614	1	4	7	58	1
2621	1	4	8	94	1
...					
...					

Berdasarkan data yang didapat pada tabel 2, terdapat perbedaan satuan yang signifikan diantara variabel-variabel yang diteliti, sehingga diperlukan proses normalisasi data. Dengan menggunakan rumus z-score, dihasilkan nilai-nilai yang sudah distandarkan.

$$Z = \frac{X - \bar{X}}{\rho}$$

Z : Z Score,
 X : nilai data
 \bar{X} : Rata-rata(mean)
 ρ : Standar deviasi

Tabel 3. Data hasil normalisasi

CustID	ZL	ZR	ZF	ZM	ZP
2596	-1.1887	-1.0030	-0.2925	-0.0942	-0.2699
2581	-1.1887	-1.0030	-0.2452	-0.0930	-0.2699
2627	-1.1887	-1.0030	-0.1979	-0.0944	-0.2699
2648	-1.1887	-1.0030	-0.1979	-0.0926	-0.1234
2620	-1.1887	-1.0030	0.1332	-0.0847	-0.2699
2638	-1.1887	-1.0030	0.4169	-0.0903	0.0231
2597	-1.1887	-1.0030	2.1668	1.9776	13.9406
2606	-1.1887	-0.6772	-0.2925	-0.0942	-0.2699
487	-1.1887	-0.6772	-0.2925	-0.0942	-0.2699
2617	-1.1887	-0.6772	0.1332	-0.0520	-0.1234
2615	-1.1887	-0.3515	-0.2925	-0.0941	-0.2699
2588	-1.1887	-0.3515	-0.1979	-0.0745	-0.2699
2616	-1.1887	-0.3515	-0.1506	-0.0915	-0.2699
2643	-1.1887	-0.0257	-0.2925	-0.0941	-0.2699
2641	-1.1887	-0.0257	-0.2452	-0.0929	-0.1234
2631	-1.1887	-0.0257	-0.2452	-0.0917	-0.2699
2610	-1.1887	-0.0257	-0.1033	-0.0807	-0.2699
2614	-1.1887	-0.0257	-0.0087	-0.0912	-0.1234
2621	-1.1887	-0.0257	0.0386	-0.0891	-0.1234
....					
.....					

Berdasarkan data yang diterima dari pihak marketing, maka didapat nilai bobot masing kriteria adalah sebagai berikut

	Length	Recency	Frequency	Monetary	Payment
Length	1	1	3	3	5
Recency		1	3	3	7
Frequency			1	1	7
Monetary				1	7
Payment					1

Gambar 1. Matrik Perbandingan AHP

	L	R	F	M	P
L	1	1	3	3	5
R	1	1	3	3	7
F	0.33	0.33	1	1	7
M	0.33	0.33	1	1	1
P	0.2	0.14	0.14	1	1
Sum	2.86	2.81	8.14	9	21
Mean	0.57	0.56	1.63	1.8	4.2

Gambar 2. Matrik Perbandingan AHP hasil aplikasi peneliti

	L	R	F	M	P	SR	EV
L	0.349	0.356	0.368	0.333	0.238	1.645	0.329
R	0.349	0.356	0.368	0.333	0.333	1.74	0.348
F	0.116	0.119	0.123	0.111	0.333	0.802	0.16
M	0.116	0.119	0.123	0.111	0.048	0.516	0.103
P	0.07	0.051	0.018	0.111	0.048	0.297	0.059

Gambar 3. Perbandingan pada LRFMP berpasangan data normalisasi hasil aplikasi peneliti

Langkah selanjutnya adalah menentukan nilai *Eigen* Maksimum dengan rumus :

$$EM = \sum SumOfCol * EV$$

$$EM = (2.86 * 0.329) + (2.81 * 0.348) + (8.14 * 0.16) + (9 * 0.103) + (21 * 0.059)$$

$$EM = 5.404$$

Menentukan nilai Indeks Konsistensi dan rasio konsistensi :

1	0	0	0.28	0.8	1.5	1.35	1.4	1.48	1.21	1.48	1.28	1.21	1.28		
0	1	5	3	4	2	0	1	8	0	10	11	15	13	14	12

Gambar 4. Daftar Indeks random konsistensi (RI)

$$CI = \frac{EM - n}{n - 1}$$

$$CI = \frac{5.404 - 5}{5 - 1}$$

$$CI = 0.101$$

$$RC = \frac{CI}{RI}$$

$$RC = \frac{0.101}{1.12} = 0.090$$

Hasil data normalisasi yang dihasilkan dikalikan dengan masing-masing bobot.

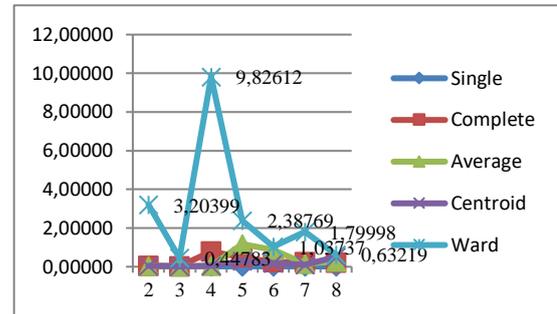
Table 4. Data Normalisasi terbobot

CustID	WL	WR	WF	WM	WP
2596	-0.3910	-0.3490	-0.0469	-0.0097	-0.0160
2581	-0.3910	-0.3490	-0.0393	-0.0096	-0.0160
2627	-0.3910	-0.3490	-0.0317	-0.0097	-0.0160
2648	-0.3910	-0.3490	-0.0317	-0.0096	-0.0073
2620	-0.3910	-0.3490	0.0214	-0.0087	-0.0160
2638	-0.3910	-0.3490	0.0669	-0.0093	0.0014
2597	-0.3910	-0.3490	0.3476	0.2043	0.8278
2606	-0.3910	-0.2357	-0.0469	-0.0097	-0.0160
487	-0.3910	-0.2357	-0.0469	-0.0097	-0.0160
2617	-0.3910	-0.2357	0.0214	-0.0054	-0.0073
2615	-0.3910	-0.1223	-0.0469	-0.0097	-0.0160
2588	-0.3910	-0.1223	-0.0317	-0.0077	-0.0160
2616	-0.3910	-0.1223	-0.0242	-0.0095	-0.0160
2643	-0.3910	-0.0089	-0.0469	-0.0097	-0.0160
2641	-0.3910	-0.0089	-0.0393	-0.0096	-0.0073
2631	-0.3910	-0.0089	-0.0393	-0.0095	-0.0160
2610	-0.3910	-0.0089	-0.0166	-0.0083	-0.0160
2614	-0.3910	-0.0089	-0.0014	-0.0094	-0.0073
2621	-0.3910	-0.0089	0.0062	-0.0092	-0.0073
....					
....					

Proses selanjutnya yaitu menganalisa semua metode *Hierarchical clustering* terhadap data normal terbobot yang didapatkan. Dan hasil *clustering* dibanding semua nilai DBI untuk mencari nilai K optimum.

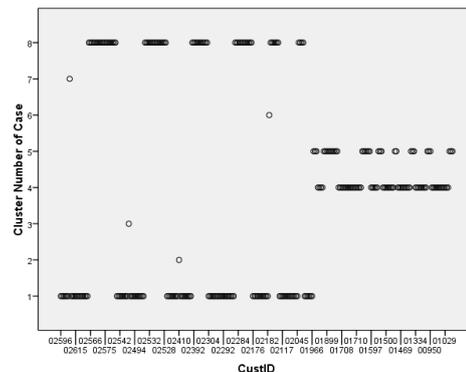
Table 5. Hasil perbandingan nilai DBI metode *Hierarchical Clustering*

Metode/K	2	3	4	5	6	7	8
Single	0.04	0.02	0.02	0.01	0.01	0.01	0.01
Complete	0.04	0.02	0.79	0.50	0.24	0.24	0.21
Average	0.04	0.02	0.05	1.17	0.87	0.16	0.27
Centroid	0.04	0.02	0.05	0.30	0.19	0.11	0.51
Ward	3.20	0.45	9.83	2.39	1.04	1.80	0.63

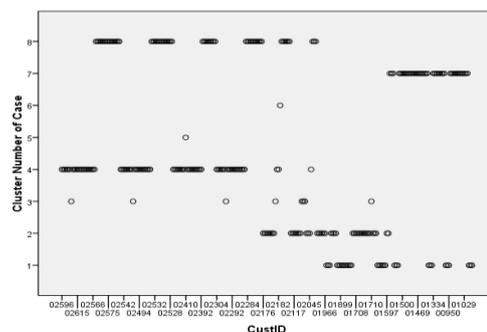


Gambar 5. Grafik perbandingan nilai DBI metode *Hierarchical Clustering*

Berdasarkan grafik diatas, maka metode yang terbaik adalah *metode single*, dengan jumlah kluster terbaik adalah 8 kluster. Setelah menemukan nilai K optimum, selanjutnya proses klustering metode *K-Mean* dengan nilai jumlah kluster adalah 8 *attribute* LRFMP dan LRFM.



Gambar 6.: Scatter Plot K-Mean LRFMP



Gambar 7. Scatter Plot K-Mean LRFM

Metode	DBI		Metode	DBI
	AHP	Non-AHP		
LRFMP	0.40839	0.56753	Single	0.00750
LRFM	1.15777	1.17851	K-Mean	0.40839

Berdasarkan hasil tersebut dapat dikatakan bahwa penambahan parameter P (*payment*) mampu membuat nilai DBI lebih baik untuk data terbobot maupun tidak. Sedangkan pada model LRFM penambahan bobot dapat meningkatkan nilai DBI *cluster*.

Berdasarkan nilai DBI hasil *clustering single* dan *k-mean* tersebut dapat diambil kesimpulan bahwa nilai K=8 hasil *clustering* metode *single* memang layak digunakan sebagai nilai awal K pada *clustering k-mean*, walaupun nilai DBI *single linkage* masih lebih baik dari K-mean.

Setelah melakukan evaluasi hasil *clustering*, langkah selanjutnya adalah melakukan analisa karakteristik tiap klaster.

Tabel 6. Nilai rata-rata attribute L, R, F, M dan P tiap kluster

K	1	2	3	4	5	6	7	8
L	3.49	3	2	11.19	9.82	5	1	2.74
R	1.96	1	1	2.03	7.61	4	1	7.54
F	7.44	305.75	88.2	7.19	3.56	0.98	53.11	2.31
M	504	50,005.22	864.74	178.88	93.15	265,856.79	34,620.41	57.56
P	1.58	-0.01	39.26	1.51	1.08	-0.01	97.66	0.98
Total	519.5	50,316.95	998.19	204.81	120.22	265,872.75	34,780.18	79.14

Tabel 7. Nilai rata-rata attribute L, R, F, M tiap kluster

K	1	2	3	4	5	6	7	8
L	9.74	7.31	4.87	2.54	3	5	12.38	2.76
R	7.78	1.42	1	2.47	1	4	2.55	7.88
F	3.6	8.13	53.86	4.19	305.75	0.98	5.37	2.34
M	62.42	271.1	8,308.84	165.94	50,005.22	265,856.79	156.24	60.8
Total	83.53	288	8,368.58	175.14	50,314.96	265,866.77	176.54	73.78

Dari data tabel 5 dan 6 dibuat domain nilai dan deskripsi setiap atribut.

Tabel 8. Deskripsi (*Linguistik*) dan Domain Nilai tiap *attribute*

Attribute	Deskripsi	Domain Nilai
Length	Baru	$1 \leq L \leq 3$
	Belum Lama	$3 < L \leq 6$
	Lama	$L > 6$
Recency	Baru	$1 \leq R \leq 4$
	Lama	$4 < R \leq 8$
	Lama Sekali	$R > 8$
Frekuensi	Jarang	$0 \leq F \leq 50$
	Agak Sering	$50 < F \leq 100$
	Sering	$F > 100$
Monetary	Rendah	$0 \leq M \leq 500$
	Sedang	$500 < M \leq 1000$
	Tinggi	> 1000
Payment	Tepat Waktu	$0 \leq P \leq 5$
	Tidak tepat waktu	$5 < P \leq 10$
	Mengabaikan	$P \geq 10$

Nilai *liguistik* pada tabel 7, dapat digunakan untuk membuat deskripsi.profile setiap kluster.

Tabel 9. Deskripsi (profile) pelanggan berdasarkan tiap kluster

K	Model	
	LRFMP	LRFM
1	Konsumen belum lama dengan frekuensi pembelian jarang, aktivitas transaksi baru, jumlah pembelian sedang dan kepatuhan pembayaran tepat waktu.	Konsumen lama dengan frekuensi pembelian lama, aktivitas transaksi lama, dan jumlah pembelian rendah.
2	Konsumen baru dengan frekuensi pembelian sering, aktivitas transaksi baru, jumlah pembelian tinggi dan kepatuhan pembayaran tepat waktu.	Konsumen lama dengan frekuensi pembelian jarang aktivitas transaksi baru, dan jumlah pembelian rendah.
3	Konsumen baru dengan frekuensi pembelian agak sering aktivitas transaksi baru, dan jumlah pembelian sedang dan mengabaikan kepatuhan pembayaran.	Konsumen belum lama dengan frekuensi pembelian agak sering, aktivitas transaksi baru, dan jumlah pembelian tinggi.
4	Konsumen baru dengan frekuensi pembelian jarang aktivitas transaksi baru, jumlah pembelian rendah dan kepatuhan pembayaran tepat waktu.	Konsumen baru dengan frekuensi pembelian jarang aktivitas transaksi baru, dan jumlah pembelian rendah.
5	Konsumen lama dengan frekuensi pembelian jarang aktivitas transaksi lama, jumlah pembelian rendah dan kepatuhan pembayaran tepat waktu.	Konsumen baru dengan frekuensi pembelian sering aktivitas transaksi baru, dan jumlah pembelian tinggi.
6	Konsumen belum lama dengan frekuensi pembelian jarang aktivitas transaksi lama, dan jumlah pembelian tinggi dan kepatuhan pembayaran tepat waktu.	Konsumen belum lama dengan frekuensi pembelian jarang aktivitas transaksi lama, dan jumlah pembelian tinggi.
7	Konsumen baru dengan frekuensi pembelian agak sering, aktivitas transaksi baru, jumlah pembelian tinggi dan mengabaikan kepatuhan pembayaran.	Konsumen lama dengan frekuensi pembelian jarang aktivitas transaksi baru, dan jumlah pembelian rendah.
8	Konsumen baru dengan frekuensi pembelian jarang aktivitas transaksi lama, jumlah pembelian rendah dan kepatuhan pembayaran tepat waktu.	Konsumen baru dengan frekuensi pembelian jarang aktivitas transaksi lama, dan jumlah pembelian rendah.

Selanjutnya hasil member *clustering* dibandingkan dengan hasil analisa (verifikasi) marketing secara manual. Berdasarkan hasil tersebut maka dapat buat table perbandingan antara jumlah member hasil verifikasi dan hasil *clustering*.

Tabel 10. Tabel Ketepatan anggota clustering model LRFMP

K	Jumlah Member		Rasio
	Y	T	
1	30	1	97%
2	41	4	91%
3	6	2	75%
4	75	4	95%

1. Metode Hierarchical Clustering Single adalah metode yang terbaik untuk penentuan nilai K awal metode K-mean dibandingkan dengan metode Single Ward, Complete Linkage, Average Linkage dan Centroid.
2. Penambahan parameter P (payment) mampu membuat nilai DBI lebih baik.
3. Penambahan bobot dapat meningkatkan nilai DBI cluster.
4. Nilai rasio akurasi model LRFM lebih baik dari model LRFMP
5. Jumlah kluster (K) hasil metode Hierarchical Clustering, bisa diterapkan untuk nilai K awal k-mean. Tetapi dengan jumlah kluster yang sama, DBI metode single linkage lebih baik dari pada metode k-mean.
6. Hasil clustering dapat dijadikan acuan marketing dalam menentukan perlakuan terhadap customer.

Daftar Pustaka

- [1] Mohammed J. Zaki and Wagner Meira Jr. 2014. *Data Mining And Analysis : Fundamental Concepts and Algorithms*, Cambridge University Press is part of the University of Cambridge
- [2] Pang-Ning Tan, Michael Steinbach dan Vipin Kumar, 2006, *Introduction to Data Mining*. Boston San Francisco New York
- [3] Jiawei Han and Micheline Kamber, 2006. *Data Mining: Concepts and Technques, Second Edition*. University of Illinois at Urbana-Champaign.
- [4] Subhash Sharma, 1996. *Applied Multivariate Techniques*.

5	1	0	100%
6	1	0	100%
7	40	2	95%
8	58	1	98%
TOTAL	252	14	95%

Rrasio akurasi keseluruhan data pada model LRFM lebih baik dari pada model LRFMP. Tetapi dari rasio akurasi per-kluster model LRFMP lebih baik.

5. Kesimpulan

- [5] University of South Carolina. New York Chichester Brisbane Toronto Singapore
- [6] A.Parvaneh, M. J. Tarokh dan H. Abbasimehr, 2014. *Combining Data Mining and Group Decision Making in Retailer Segmentation Based on LRFMP Variables*. IUST Publication ,JIIEPR, Vol. 25, No. 3, All Rights Reserved
- [7] Rachid Ait Daoud, Abdellah Amine, Belaid Bouikhalene, Rachid Lbibb 205, *Customer Segmentation Model in E-commerce Using Clustering Techniques and LRFM Model: The Case of Online Stores in Morocco*. World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:9, No:8, 2015
- [8] Duen-Ren Liu, Ya-Yueh Shih, 2004 *Integrating AHP and data mining for product recommendation based on customer lifetime value*. Elsevier B.V. All rights reserved.
- [9] Yohana Nugraheni, 2011. *Data Mining Dengan Metode Fuzzy Untuk Customer Relationship Management (CRM) Pada Perusahaan Retail*. Tesis Universitas Udayana Denpasar.
- [10] Yani Soraya, 2011. *Perbandingan Kinerja Metode Single Linkage, Metode Complete Linkage dan Metode K-means dalam Analisis Cluster*, Skripsi Universitas Negeri Semarang
- [11] Bilson Simamora, 2005. *Analisis Multivariat Pemasaran*. PT Gramedia Pustaka Umum
- [12] Thomas L. Saaty, 2008. *Decision Making With The Analytic Hierarchy Process*. University of Pittsburgh.