

## Segmentasi Pelanggan Internet Service Provider (ISP) Berbasis Pillar K-Means

Abd Hadi

STMIK Asia Malang

e-mail: hadi@asia.ac.id

**ABSTRAK** Perusahaan penyedia layanan Internet Service Provider (ISP) memiliki jumlah pelanggan yang sangat banyak dan beragam. Dengan semakin banyak dan beragamnya jumlah pelanggan perusahaan akan sulit untuk mengetahui tipe pelanggan yang dimiliki oleh perusahaan. Akibatnya perusahaan akan kesulitan menerapkan strategi pemasaran yang tepat kepada konsumen. Dalam paper ini digunakan metode pillar K-means untuk melakukan segmentasi pelanggan. Algoritma pillar K-means untuk melakukan segmentasi pelanggan. Algoritma pillar merupakan metode optimasi untuk menentukan *centroid* awal dalam algoritma K-Means. Dengan mengoptimasi *centroid* awal maka akan menghasilkan *cluster* yang lebih baik. Setelah memperoleh hasil *cluster* yang optimal selanjutnya tipe pelanggaran dianalisis dengan menggunakan metode RFM (*Recency, Frequency, Monetary*). Hasil penelitian ini menunjukkan bahwa pillar K-means mampu mengoptimasi hasil *cluster* :  $k = 4$  dengan  $a = 0.5$  dan  $b = 0.8$  serta nilai *silhouette*  $8 = 0.47103$ . Dari hasil segmentasi 150 pelanggan diperoleh tipe pelanggaran yang terdiri *Most Valuable Customers* (33) *Most Growable Customers* (41), *Migrators* (23) dan *Below Zero* (53).

**Kata Kunci:** *segmentasi pelanggan, pillar K-Means, RFM*

**ABSTRACT.** Internet service provider (ISP) service companies have a very large and diverse number of customers. With more and more number of company customers, it will be difficult to know the type of customer's company. As a result the company will have difficulty implementing the right marketing strategy to consumers. In this paper the K-means pillar method is used to segment customers. Pillar K-means algorithm for segmenting violations. The Pillar algorithm is an optimization method for determining initial centroid in the K-Means algorithm. By optimizing the initial centroid it will produce a better cluster. After obtaining the optimal cluster results, the types of violations are analyzed using the RFM (*Recency, Frequency, Monetary*) method. The results of this study indicate that the K-means pillars are able to optimize cluster results:  $k = 4$  with  $a = 0.5$  and  $b = 0.8$  and *silhouette* value  $8 = 0.47103$ . From the results of segmenting 150 subscribers, the type of subscription consists of *Most Valuable Costmers* (33) *Most Growable Costomers* (41), *Migrators* (23) and *Below Zero* (53).

**Keywords:** *customer segmentation, pillar K-Means, RFM*

### 1. PENDAHULUAN

Saat ini data mining umumnya digunakan oleh perusahaan dengan jumlah pelanggan yang besar seperti perusahaan finansial, telekomunikasi, dan organisasi marketing (Neethu 2012). Banyaknya jumlah pelanggan menuntut perusahaan memperhatikan *customer relationship management* (CRM). CM membantu membangun hubungan jangka panjang dengan pelanggan (Ling R dan Yen 2001). Factor kunci dalam pengembangan strategi CRM adalah memahami dan menganalisis perilaku pelanggan (Ngai dan Li Xiu 2009). PT Telkom sebagai salah satu perusahaan penyedia layanan telekomunikasi di Indonesia tentunya memiliki data pelanggan yang sangat besar, sehingga pengelolaannya memerlukan data mining. Data mining (DM) dapat digunakan oleh perusahaan untuk penentuan keputusan, peramalan, dan prediksi perilaku pelanggan di masa depan (Ngai dan Li Xiu 2009). Dengan menerapkan konsep *analytical* CRM dalam data mining, perusahaan dapat melakukan identifikasi pelanggan dengan melakukan segmentasi pelanggan. Segmentasi pelanggan dilakukan untuk mengetahui perilaku dan tipe pelanggan. Membentuk kedalam kelompok yang berbeda membantu meningkatkan kualitas rekomendasi, membantu perusahaan dalam membuat keputusan dengan mengidentifikasi pelanggan secara lebih jelas untuk mengembangkan strategi pemasaran yang lebih baik (Derya 2011). Tujuan utamanya adalah untuk mendatangkan keuntungan bagi pihak perusahaan.

Teknik yang dapat digunakan untuk segmentasi pelanggan adalah *clustering* merupakan pengelompokan sejumlah data atau objek ke dalam *cluster* yang sama (Kohe aria 2007). Metode yang paling umum digunakan dalam *clustering* adalah algoritma *K-means*. *K-means* merupakan metode *clustering* yang populer karena, *K-means* mempunyai kemampuan mengelompokan data dalam jumlah yang besar dan mampu mendeteksi *outlier* dengan cepat dan efisien (Kohe aria 2007).

Permasalahan utama dalam metode *K-means* adalah penentuan pusat *cluster* yang dibandingkan secara acak. Karena penentuan titik awal dilakukan secara acak, *k-means* tidak menjamin hasil *clustering* yang unik (Shehroz 2004). Untuk dapat mengetahui perilaku dan tipe konsumen, perlu dilakukan optimasi pada algoritma *K-means* sehingga menghasilkan *cluster* yang optimal.

Ada beberapa metode yang dapat digunakan untuk mengatasi permasalahan inisialisasi *cluster* dalam *K-Means*. Bradley dan Fayyad (1988) mengusulkan sebuah algoritma yang memurnikan *initial points* dengan menganalisis distribusi data dan probabilitas kerapatan data (Bradley dan Fayyad 1988). Barakbah dan Helen (2005) membahas algoritma baru, yang disebut *optimized K-means*, ide utamanya adalah menyebarkan *centroid* awal dalam ruang fitur sehingga jarak di antara mereka menjadi sejauh mungkin (Barakbah dan Helen 2005). Barakbah dkk. (2005) menyajikan optimalisasi titik awal pada *K-Means* dengan menggunakan *simulated annealing* (Barakbah 2005). Barakbah (2006) mengusulkan algoritma baru untuk mengoptimalkan *centroid* awal pada penggunaan *K-means* dengan cara memisah penempatan *centroid* awal sejauh mungkin dalam pendistribusian data. Dengan mempertimbangkan perhitungan optimasi pada *K-means*. Barakbah dan Arai (2007) menyajikan algoritma untuk menentukan *centroid* awal dengan *hierarchical K-means* (Kohe aria 2007). Barakbah dan Kiyoki (2009) menyajikan pendekatan baru untuk optimasi *K-means* dengan perhitungan *Distance Maximization* untuk penentuan *centroid* awal yang disebut algoritma *Pillar*.

Berdasarkan penelitian terkait optimasi *K-means* yang telah dijabarkan maka dalam penelitian ini akan menggunakan *PillarK-means* untuk optimasi hasil dari *segmentasi* data pelanggan. Tujuan dari penelitian yang akan dilakukan adalah membentuk *segmentasi* data pelanggan dengan menggunakan Algoritma *pillar K-means* dan metode *RFM* untuk mengetahui tipe pelanggan dari PT. Telkom. Algoritma *Pillar* mampu mengoptimalkan pemilihan *centroid* awal dan meningkatkan ketepatan pada proses *segmentasi* di algoritma *K-means*. Algoritma *pillar* juga mampu menangani *outlier* dengan mekanisme *outlier detection*. Selain itu waktu eksekusi (*execution time*) algoritma *pillar* juga menunjukkan kinerja yang lebih baik daripada algoritma optimasi *centroid* awal lainnya (Barakbah dan Kiyoki 2009).

Pembahasan terkait paper ini akan dijabarkan sebagai berikut: *Section 2* membahas tentang penelitian terkait. *Section 3* membahas mengenai methodology yang meliputi deskripsi data. Kerangka penelitian, data *preprocessing*, analisis *RFM*, algoritma *K-means*, algoritma *pillar* dan evaluasi hasil *cluster*. Pada *Section 4* akan dibahas mengenai hasil penelitian dan pembahasan. Kesimpulan akan dibahas pada *Section 5*.

## 2. PENELITIAN TERKAIT

Penelitian di bidang *segmentasi* pelanggan telah banyak dilakukan sebelumnya, ada berbagai metode yang digunakan dalam *segmentasi* pelanggan seperti misalnya dengan menggunakan SOM (*self organizing map*) (Rusmeier 1997), *K-means* (Qioru cha 2012), *K-medoids* (Han Zho 2014), *Association Rule Mining* (Gulluoglu 2015). Selain itu juga penelitian terkait *segmentasi* pelanggan mengkombinasikan dua metode atau lebih seperti *Density Based DBSCAN* dan *K-means* untuk *segmentasi* data perbankan (Zakrzawka 2005). Mengkombinasikan PCA (*Principal Component Analysis*) dengan *Back Propagation Neural Network* untuk menganalisis perilaku konsumen (Han Minghua 2008). Seluruh penelitian yang dilakukan mampu membentuk *segmentasi* pelanggan dengan berbagai jenis studi kasus. Untuk penelitian pada bidang *segmentasi* pelanggan dengan menggunakan *Pillar K-means* belum pernah dilakukan sebelumnya.

## 3. METODE PENELITIAN

### A. Deskripsi Data

Data yang digunakan dalam penelitian ini adalah data transaksional PT. Telkom Indonesia pada tahun 2004. Data transaksional tersebut terdiri dari data *Divisi Enterprise Service* (DES) dan *Divisi Business Service* (DBS). Data yang digunakan terdiri dari beberapa atribut seperti : ID nasabah, nama wilayah, produk, sub produk, ID wilayah, nama wilayah dan *revenue*. Setiap *record* data berisi catatan periode waktu transaksi pelanggan. Periode waktu yang digunakan adalah satu bulan, dan berlangsung selama periode bulan Januari – Juli tahun 2014.

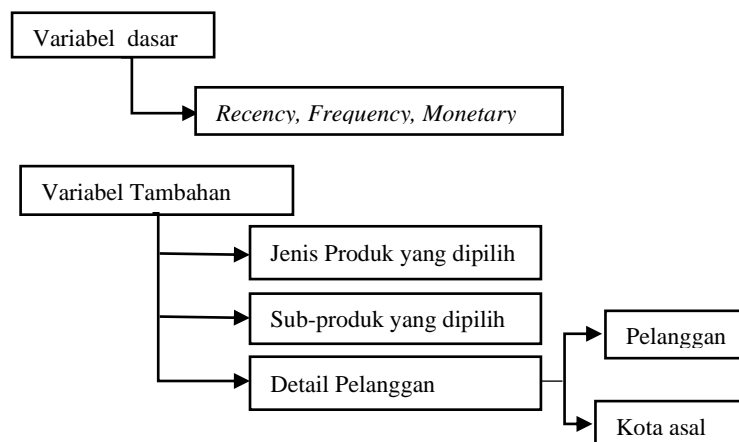
### B. Kerangka Penelitian

Penelitian dimulai dengan melakukan pengumpulan data dan informasi. Selanjutnya dilakukan tahapan *preprocessing* dilakukan analisis *RFM*. Pada waktu *segmentasi* dilakukan proses *clustering* hasil *cluster*. Terakhir tipe pelanggan ditentukan dari hasil *segmentasi* analisis *RFM*. Kerangka penelitian dapat digambarkan sesuai **Gambar 1**. Program yang digunakan untuk analisis data adalah *Python 2.7.10*, dengan editornya menggunakan *Sublime Text 3*.

### C. Preprocessing Data

Sebelum dapat analisis, perlu dilakukan preprocessing terhadap data mentah yang diperoleh dan dijabarkan pada deskripsi data. Preprocessing data merupakan langkah yang diperlukan untuk membuat proses *knowledge discovery* menjadi lebih mudah dan tepat. Data yang digunakan adalah data pelanggan dan data historikal transaksi penjualan PT. Telkom untuk periode Januari – Juli 2014. Data mentah yang diperoleh dari data pelanggan dan data historikal penjualan tersebut akan melalui praproses data yang meliputi : *data integration, data cleaning, data transformation* dan *data selection*.

*Data integration* merupakan proses untuk mengintegrasikan data yang berasal dari *database* yang berbeda. *Data cleaning* dilakukan untuk menangani data yang memiliki *missing value*. *Data transformation* bertujuan untuk merubah data menjadi format yang lebih sesuai. *Data selection* merupakan proses untuk memilih atribut yang akan digunakan dalam analisis. Proses *data selection* dilakukan dengan menggunakan metode RFM. Setelah dilakukan preprocessing data maka variabel yang digunakan dapat dilihat pada Gambar 2.



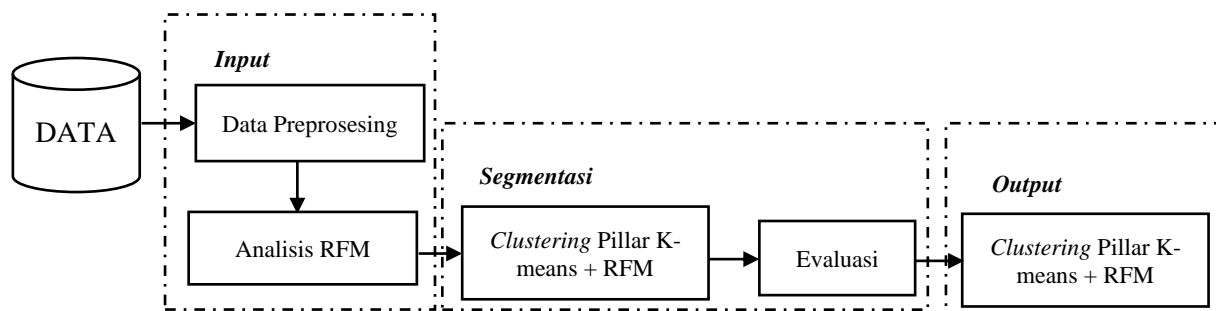
**Gambar 1.** Identifikasi Variabel hasil Preprocessing

### D. Analisis RFM

Konsep RFM diperkenalkan oleh Bult dan Wansbeek dan (1995) dan telah terbukti sangat efektif bila diterapkan pada database marketing (Blattbeng 2008). Analisis RFM tergantung pada *Recency* (R), *Frequency* (F), dan *Monetary* (M). Ketiga komponen tersebut mempengaruhi kemungkinan transaksi terakhir yang dilakukan pelanggan. *Recency*, merupakan kriteria yang mengacu interval rentang waktu transaksi terakhir pelanggan dengan periode analisis. *Frequency*, mengacu pada jumlah transaksi dalam periode tertentu, misalnya hari, bulan atau tahun. *Monetary*, mengacu pada jumlah total transaksi dalam periode tertentu Analisis RFM digunakan untuk *Scaling* atribut  $R - F - M$ .

Langkah – langkah analisis RFM dapat dijabarkan sebagai berikut:

1. Urutkan data atribut R-F-M secara *descending* atau *ascending*.
2. Bagi ketiga atribut R-F-M kedalam 5 bagian, dan setiap bagian terdiri dari 20% nilai total. Kelima bagian akan diberi nilai 5, 4, 3, 2, 1, setiap angka tersebut. Mencerminkan kontribusi tertinggi pelanggan. Nilai 5 mengacu pada pelanggan dengan kontribusi tertinggi, nilai 1 mengacu pada pelanggan dengan kontribusi terendah pada *revenue*.
3. Ulangi langkah 1 dan 2 untuk setiap atribut R-F-M. Total 125 kombinasi (5 x 5 x 5) dengan atribut R-F-M berskala 5 (5, 4, 3, 2, dan 1).



**Gambar 2** Kerangka Penelitian Analisis RFM dan Segmentasi

**E. Algoritma K-means**

Algoritma *K-means* dapat dijabarkan dengan proses untuk menentukan pemetaan  $f: D \rightarrow C$  dari beberapa data  $D = \{d_1, d_2, \dots, d_n\}$  dengan beberapa *cluster*  $C = \{c_1, c_2, \dots, c_n\}$  berdasarkan similarity antara  $d_i$ . Misalkan  $A = \{a_i | i = 1, \dots, f\}$  menjadi atribut vektor  $f$ -dimensi dan  $X = \{x_i | i = 1, \dots, k\}$  dimana  $M_{\epsilon} X$  adalah  $M_i = \{m_{ij} | j = 1, \dots, n(s_i)\}$  sebagai anggota  $s_i$ , dimana  $n\{s_i\}$  adalah jumlah anggota untuk  $\{s_i\}$ . Setiap *cluster* memiliki pusat *cluster* (*centroid*) dari tiap element segmentasi. Algoritma *clustering* dapat dijabarkan sebagai berikut (Barakbah dan Kiyoki 2009).

1. Memulai algoritma dengan menentukan *starting point centroid* awal  $C$  secara acak.
2. Menghitung jarak  $d$  antara  $X$  kepusat *cluster*  $C$ . Untuk menghitung jarak dapat menggunakan *Euclidean distance*.

$$\delta(x - y) = \|x - y\|_2 = \left( \sum_{i=1}^d (x - y)^2 \right)^{1/2} \tag{1}$$

3. Pisahkan  $x_i$  untuk  $\tau = 1, \dots, n$  ke  $S$  yang memiliki nilai minimum  $\delta(x_i, C)$ .
4. Tentukan pusat *cluster* baru  $C_i$  untuk  $i = 1, \dots, k$ .

$$C_i = \frac{1}{n_i} \sum_{j=1}^{n(s_i)} m_{ij} \in S_i \tag{2}$$

5. Kembali ke langkah 2, sampai semua *centroid convergent*.

*Centroid* dapat dikatakan *Convergent* jika posisinya tidak berubah lagi selama proses iterasi. Perubahan posisi *centroid* juga dapat berhenti pada iterasi ke  $\tau$  dengan *threshold*  $\epsilon$  (Kovesi 2011), jika posisi tersebut telah diperbaharui oleh jarak yang lebih kecil  $\epsilon$ .

$$\left| \frac{C^1 - C^{t-1}}{C^1} \right| < \epsilon \tag{3}$$

Optimasi penentuan *centroid* awal untuk algoritma *K-means* dapat dilakukan dengan menggunakan algoritma *Pillar*. Dalam penelitian ini algoritma *Pillar* digunakan sebagai penentu jumlah *cluster* optimal serta digunakan untuk menentukan *centroid* awal untuk masing-masing *cluster*. Selanjutnya jumlah *cluster* optimal serta hasil perhitungan *centroid* awal pada algoritma *Pillar* digunakan pada algoritma *K-means*. Karena menggunakan *centroid* awal dari perhitungan algoritma *Pillar* maka penentuan *centroid* awal tidak lagi ditentukan secara acak.

**F. Algoritma pillar**

Algoritma *pillar* terinspirasi dari proses berpikir untuk menentukan satu set lokasi *pillar* untuk membuat bangunan atau rumah yang stabil. Dengan cara mendistribusikan posisi *pillar* sejauh mungkin antar satu sam lain pilar dapat menahan tekanan atap atau menstabilkan bangunan atau rumah. Distribusi *pillar* dianggap sebagai *centroid* sedangkan tekanan gravitasi dari atap dianggap sebagai distribusi data dalam sebuah ruangan vektor. Dalam algoritma *Pillar* penentuan posisi *centroid* awal berdasarkan akumulasi jarak terjauh *centroid* dalam distribusi data (Barakbah dan Kiyoki 2009).

Algoritma *Pillar* dijabarkan sebagai berikut. Misalnya  $X = \{x_i | i = 1, \dots, n\}$  merupakan data - data,  $k$  adalah jumlah *cluster*,  $C = \{C_i | i = 1, \dots, k\}$  merupakan inisialisasi *centroid*,  $SX \subseteq X$  akan diidentifikasi untuk  $X$  yang siap dipilih pada saat melakukan urutan proses.  $DM = \{x_i | i = 1, \dots, n\}$  adalah akumulasi *distance* matrik,  $D = \{x_i | i = 1, \dots, n\}$  adalah *distance* matrik untuk iterasi selanjutnya (Barakbah dan Kiyoki 2009). Tahapan algoritma *Pillar* adalah sebagai berikut:

1. Set  $C = \emptyset, SX = \emptyset$  dan  $DM = []$
2. Hitung  $D \leftarrow dis(X, m)$
3. Tentukan jumlah ketetanggaan  $N_{min} = \alpha \cdot n/k$
4. Tentukan nilai  $d_{max} \leftarrow ar_{max}(D)$
5. Tentukan batas ketetanggaan  $nb_{dis} = \beta \cdot d_{max}$
6. Tetapkan  $i = 1$  sebagai *counter* untuk menentukan *centroid* awal ke -  $i$
7.  $DM = DM + D$
8. Pilih  $x \leftarrow x_{argmax}(DM)$  sebagai kandidat untuk inisialisasi *centroid* ke -  $i$
9.  $SX = SX \cup x$
10. Tetapkan  $D$  sebagai jarak antara matrik  $X$  dan
11. Tetapkan  $no \leftarrow number$  data yang memenuhi  $D \leq nb_{dis}$
12. Tentukan  $DM(x) = 0$
13. Jika  $no \leftarrow n_{min}$ , kembali ke langkah 8
14. Tentukan  $D(SX) = 0$
15.  $C = C \cup x$
16.  $i = i + 1$
17. Jika  $i \leq k$ , kembali ke langkah 7
18. selesai,  $C$  adalah solusi *centroid* awal yang optimal

#### G. Pengujian Hasil Clustering

Untuk mengetahui kualitas dari *cluster* yang dihasilkan dari proses *clustering* dengan menggunakan *pillarK-means*, maka *cluster* yang dihasilkan perlu diuji terlebih dahulu. Pengujian kualitas *cluster* dapat dilakukan dengan menggunakan fungsi *silhouette* (Peter J 1987). Fungsi *silhouette* menunjukkan apakah sebuah objek sudah tepat diletakkan pada suatu *cluster* atau tidak. *Silhouette value*  $s = (i)$  dari objek  $i$  dapat dijabarkan pada persamaan (4). Nilai dari  $s = (i)$  berada pada rentang  $-1$  dan  $1$   $s \in [-1,00,1,00]$ .

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4)$$

Selain dengan menggunakan fungsi *Silhouette*, pengujian juga dapat dilakukan dengan menggunakan *SSE (Sum squared Error)* seperti yang dilakukan pada persamaan (5). Dimana  $k$  adalah jumlah *cluster* dan  $C_i$  merupakan pusat *cluster* ke -  $i$ .

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} dist(c_i, x)^2 \quad (5)$$

### 3. HASIL DAN PEMBAHASAN

#### A. Hasil preprosesing dan analisis RFM

Dalam penelitian ini digunakan 150 record data pelanggan. Sebelum dilakukan analisis RFM dilakukan proses preprosesing yang terdiri dari *data integration*, *data cleaning*, *data information* dan *data selection*. Untuk mempermudah analisis dilakukan *scaling* untuk masing masing atribut RFM. Hasil preproses data pelanggan disajikan pada Tabel 1. Proses *scaling* dengan skala 5 dilakukan dengan rentang 20% dari masing masing data, seperti pada tabel 2. Hasil *scaling* data dengan menggunakan analisis RFM disajikan pada tabel 3.

**Tabel 1.** Data pelanggan

ID	Kota	Produk	Sub produk	Recency	Frequency	Monetary
B60	Jakarta utara	Pend content speedy	Legacy,wireline,isdn,dome stic call,wline	Juli-2014	19	12.400.000
B61	Bogor	Pend abonement speedy multi speed	New wave,wip,usage wline,mouthly bee wline	Juni-2014	12	6.130.000
B62	Bandung	Pend usage telkownet intan	Legacy,data communication,usage wline,mouthly bee wline	April-2014	6	2.579.00
B63	Jakarta pusat	Pend content speedy	Legacy,data communication,usage,wline,mothly bee wline	Juni-2014	21	13.700.000

**Tabel 2.** Scaling data RFM

Scaling	Reency (bulan)	Frequency	Monetary
5	7	>20	>12.000.000
4	6	16-20	9.000.001 - 12.000.000
3	4-5	11-15	6.000.001 - 9.000.000
2	2-3	6-10	3.000.001 – 6.000.000
1	0-1	0-5	0 – 3.000.000

**Tabel 3.** Hasil *scaling* data

ID	Recency	Frequency	Monetary	RFM
060	5	4	5	545
061	4	3	4	434
062	3	2	1	321
063	5	5	5	555

**Tabel 4.** Jumlah cluster terbaik untuk setiap *k*, berdasarkan nilai silhouette.

K	$\alpha$	$\beta$	Nilai silhouette	Jumlah cluster kosong
2	0.85	0.4	0.487646509	0
3	0.3	0.4	0.569886169	0
4	0.25	0.3	0.617346845	0
5	0.3	0.3	0.561355816	0
6	0.3	0.3	0.658983411	0
7	0.4	0.3	0.80495978	0
8	0.2	0.25	100000000	0
9	0.3	0.1	100000000	
10			100000000	

**Tabel 5.** Solusi cluster terbaik berdasarkan hasil analisis dan perhitungan SSE.

A	$\beta$	Negative avg. silhouette	Cluster kosong	SSE
0.45	0.75	0	0	3779.556
0.5	0.8	0	0	3464.366
0.55	0.75	15	0	3626.871
0.65	0.8	15	0	3626.871
0.7	0.75	16	0	3633.169

**Tabel 6.** Penentuan tipe pelanggan

Cluster	Ukuran	Avg.recency	Avg.frequency	Avg.monetary	Pola RFM	Tipe pelanggan
C1	33	4.727	4,758	4.636	R↑F↑M↑	Most valuable customer
C2	41	3.487	3.927	4.512	R↓F↑M↑	Most growable customer
C3	23	3.826	1.696	4.130	R↓F↓M↑	Migratos
C4	53	3.712	4.000	2.308	R↓F↑M↓	Below zero
total	150	3.938	3.595	3.897		

**B. Hasil segmentasi dan pengujian hasil segmentasi**

Dari hasil percobaan yang telah dilakukan ditemukan bahwa nilai  $\alpha$  dan  $\beta$  memiliki peran yang signifikan dalam menentukan nilai silhouette. Percobaan dilakukan dengan jumlah cluster  $2 \leq K \leq 10$ . Diperoleh 178 hasil kombinasi nilai  $\alpha$  dan  $\beta$  untuk masing masing  $K$  cluster berbeda. Karena banyaknya jumlah kombinasi maka hanya akan dianalisa hasil *cluster* dengan nilai silhouette tertinggi dari masing masing *cluster*. Nilai silhouette dapat terlihat pada tabel 4.

Pada tabel 4 terlihat bahwa ada 3 nilai silhouette tertinggi pada jumlah cluster 8, 9 dan 10. Namun apabila diamati, pada cluster 9 dan 10 memiliki jumlah cluster kosong. *Cluster* kosong artinya ada cluster tanpa ada anggota cluster didalamnya. Berdasarkan hal tersebut maka ada 3 kondisi tambahan yaitu digunakan untuk menentukan jumlah cluster optimal yaitu:

1. Jika tidak memiliki cluster kosong
2. Tidak memiliki nilai average silhouette yang negative. Hal ini karena nilai silhouette terbebar belum tentu menunjukkan solusi cluster terbaik.
3. Dengan menentukan nilai SSE (*sum square error*) terkecil.

Hasil pengujian menunjukkan bahwa cluster terbaik adalah  $K = 4$  dengan  $\alpha = 0.5$  dan  $\beta = 0.8$  serta nilai silhouette  $S = 0.47103$ , dan memenuhi ketiga kriteria pengujian yang ada. Hasil pengujian selengkapnya disajikan pada tabel 5.

**C. Penentuan tipe pelanggan**

Hasil segmentasi pelanggan dengan menggunakan pillar K-means menghasilkan empat cluster yang ditentukan berdasarkan transaksi terakhir (*Recency*), frekuensi pembelian (*Frequency*) dan total nilai pembelian (*Monetary*). Kombinasi input dari nilai RFM dapat ditentukan berdasarkan tanda ↓ atau ↑. Penentuan tanda tersebut berdasarkan rata rata nilai RFM, kemudian dibandingkan dengan total rata rata nilai RFM. apabila nilai rata rata RFM lebih besar dari total rata rata nilai RFM

maka diberi tanda  $\uparrow$ , sebaliknya jika lebih kecil akan diberi tanda  $\downarrow$ . tipe pelanggan yang dapat dibentuk adalah sebagai berikut:

- 1) Most valuable customer (MVC), merupakan kelompok pelanggan dengan customer value tinggi, pola RFM nya adalah  $R \uparrow F \uparrow M \uparrow$
- 2) Most growable customer, merupakan kelompok pelanggan dengan potensi tinggi, pola RFMnya adalah  $R \downarrow F \uparrow M \uparrow$
- 3) Migrator, merupakan kelompok pelanggan yang berbeda di antara below zeros dan most growable customer, pola RFM nya adalah  $R \downarrow F \downarrow M \uparrow$
- 4) Below zeros, merupakan kelompok pelanggan dengan value yang rendah, pola RFMnya adalah  $R \downarrow F \uparrow M \uparrow$ .

Hasil penentuan tipe pelanggan berdasarkan analisis RFM disajikan pada Tabel 6.

#### 4. KESIMPULAN DAN SARAN

Hasil *segmentasi* pelanggan dengan menggunakan algoritma pillar *K-Means* menghasilkan jumlah *cluster optimum*  $K=4$  Dengan  $\alpha = 0.5$  dan  $\beta = 0.8$  serta nilai *silhouette*  $S = 0.47103$ . untuk menentukan kualitas hasil *segmentasi* tidak dapat hanya menggunakan satu parameter, melainkan dapat menggunakan 3 parameter yaitu nilai *silhouette*, *negative average silhouette* dan *sum square error (SEE)*. Dari hasil *segmentasi* pelanggan diperoleh tipe pelanggan dengan rincian *most valuable customers* sebanyak 33 pelanggan, *most growable customers* sebanyak 41 pelanggan, *migrators* sebanyak 23 pelanggan dan *below zero* sebanyak 53 pelanggan.

#### DAFTAR RUJUKAN

- Neethu Baby and Priyanka L.T *customer classification and prediction based on data mining technique. International journal of emerging technology and advanced engineering*, 2(12):314-318, December 2012.
- R Ling and Yen D. *Customer relationship management: An analysis framework and implementation strategies. Journal of computer information Systems*, 41:82-97, 2001.
- E.W.T. Ngai, Li Xiu, and D.C.K. Chau. *Application of data mining techniques in customer relationship management: A Literature review and classification. Science direct expert Systems with applications*, 36:2592-2602, 2009
- Derya, Birant. *Data Mining Using RFM Analysis, knowledge oriented applications in data mining*, chapter 6 pages 91-108. In Tech DOI: 10.5772/13683. Available <http://www.intechopen.com/books/knowledge-oriented-applications-in-data-mining/data-mining-using-rfm-anaysis>, 2011.
- Kohei Arai and A. R. Barakbah. *Hierarchical k-means; an algorithm for cendroids initialization for k-means. Reports of the faculty of science and engineering, saga university*, 36(1); 25-31, 2007.
- Shehroz S. Khan and Amir Ahmad. *Cluster center initialization algorithm for k-means clustering. Science Direct*, 15;1293-1302, 2004.
- P.S Bradley and U.M Fayyad. *Refining initial points for k-means clustering. Proc. 15<sup>th</sup> internat. Conf. on Machine Learning (ICMLS98)*, 1998
- A.R Barakbah and Amir Ahmad. *Cluster center initialization algorithm of initial centroids optimization for k-means. In proc. Soft Computing, intelligent system, and information technology (SIIT) 2005*, number 2-63-66. Petra Christian University , 2005
- A.R Barakbah, A. Fariza, and Y. Setiowati. *Optimization of initial centroids for k-means using simulated annealing. In Proc. Industrial Electronics seminar (IES) 2005*, pages 286-289. Electronic Engineering Polytechnic Institute of Surabaya-ITS, 2005.
- A.R Barakbah. *A new algorithm for optimization of k-means clustering with determining maximum distance between centroids. In proc. Industrial electronics seminar (IES) 2006*, number 240-244. Electronic Engineering Polytechnic Institute of Surabaya-ITS, 2006
- A.R Barakbah and Y. Kiyoki. *A pillar algorithm for k-means optimization by distance maximization for initial centroid designation. In Symposium on*, pages 61-68, March 2009.
- H. Rushmeir, R. Lawrence, and G. Almasi. *Case study: visualizing customer segmentation based on cluster analysis. In computer Science and Information Processing (ICMSE), 2012 International Conference on*, pages 1189-1182, 2012



- Cai Qiuru, Lou Ye, Xi Haixu, Liu Yijun, And Zhu Guangping. Telecom costumer segmentation based on cluster analysis. In *Computer Science and Information Processing (CSIP), 2012 International Conference on*, pages 103-109, 2019
- Zhao Han, Zhang Xiao –Hang, Wang Qi ,Zhang Ze-cong, and Wang Cen-Yue. Customer segmentation on mobile online behaviour. In *management science & engineering (ICMSE), 2014 international conference on, pages 103-109,2014.*
- S.S Sulluoglu. Segmenting customers with data mining techniques.in digital information, networking, and wireless communications (DINWC),2015 *Thid Internatinal Conferences on*, pages 154-159,2015
- D.Zakrzewska and J. Murlewski. Clustering algorithms for bank customer segmentation. In *intelligent systems design and applications, 2005.ISDA '05. Proceedings. 5<sup>th</sup> INTERNATIONAL CONFERENCE ON PAGES 197-202,2005*
- Minghua Han. Customer segmentation model based on retail consumer behaviour analysis. In intelligent information technology application workshops, 2008. IITAW '08. International symposium on, pages 914-917, 2008.
- R.C Blattberg, Kim B-D., and S.A. Neslin. Database Marketing: analyzing and managing customers. Springer, 2008.
- B.Kovesi, J.M.Boucher,and S.Saoudi. Stochastic k-means algorithm for vector quantization. *Pattern recognition lett*, 22:603-610,2011.
- Peter J. Rousseeuw. Silhouettes: graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53-65, 1987.t