

SISTEM PREDIKSI *CUSTOMER LOYALTY* DENGAN METODE SEGMENTASI RFM DAN PENGELOMPOKKAN DENGAN ALGORITMA *CLUSTERING MODIFIED K-MEANS* PADA PERUSAHAAN PU IBUKOTA

Herwin¹⁾Teny Handhayani²⁾Dyah E Herwindiati³⁾

Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara
Jl. Let. Jend. S. Parman No. 1, Jakarta 11440 Indonesia
email : 535110011@fti.untar.ac.id, tenyh@fti.untar.ac.id, dyahh@fti.untar.ac.id

ABSTRACT

K-Means Clustering algorithm is an algorithm that is used to classify the data that has value distance between adjacent objects in one cluster and value between object from other clusters. Centroid value on k-means clustering taken from random numbers, so that when determining the centroid will be modified so that the clustering results will not be unchanged. By using a modified algorithm k-means clustering, data customer can be grouped consistently and data on the cluster will not be changed. Measurement used to assess the customer has a high loyalty to the company or not is by using the RFM segmentation. RFM is a ranking system which is used to assess whether the customer has a high value of loyalty to the company or not.

The goal of themaking an application that the company can find customers who have a high loyalty to companies that serve as consideration for the company to be given a discounted price.

The test results stated that the results are optimal cluster and have a high RFM value contained in the third cluster. Cluster show that 64% of customer data contained in the cluster, in fact has been given a discount by the company.

Key words

RFM, Clustering, K-Means, Evaluation Silhouette, dan Evaluation Variance.

1. Pendahuluan

Masalah perusahaan yang dihadapi adalah sulitnya mengetahui pelanggan yang memiliki nilai loyalitas yang tinggi terhadap perusahaan. Penyelesaian terhadap masalah yang dihadapi perusahaan adalah dengan membuat Aplikasi Prediksi *Customer Loyalty*. Aplikasi Prediksi *Customer Loyalty* merupakan aplikasi yang dibuat dengan menggunakan segmentasi RFM dan *Clustering Modified K-Means*. Studi kasus yang

digunakan dalam aplikasi adalah data pelanggan pada perusahaan PU IBUKOTA periode dua tahun Januari 2013 – Desember 2014 dengan data pelanggan sebanyak 66 data dan data transaksi sebanyak 1156 data. Segmentasi RFM digunakan sebagai tolak ukur nilai loyalitas pelanggan terhadap perusahaan.

Pengelolaan data pelanggan dan transaksi pelanggan dilakukan dengan memindahkan data perusahaan kedalam basis data. Data yang telah disimpan dalam basis data akan diproses untuk dilakukan proses *clustering* sehingga data dapat terkelompok dalam satu kluster sesuai dengan input-an kluster.

Clustering Modified K-Means digunakan untuk mengelompokkan data yang memiliki jarak nilai yang berdekatan dalam satu *cluster* secara konsisten dengan nilai *centroid* yang tidak berubah. Modul dan unjuk kerja program telah dilakukan pemeriksaan dan dapat disimpulkan bahwa aplikasi sudah dapat berjalan dengan baik sesuai dengan hasil yang diharapkan.

Evaluasi yang digunakan untuk menguji kualitas kluster menggunakan evaluasi *variance* dan evaluasi *global silhouette*. Gambar grafik *chart* digunakan untuk membandingkan nilai RFM antar kluster dan melihat nilai RFM secara rinci. Gambar *scatter plot* digunakan untuk melihat klusterisasi apakah data yang terdapat pada satu kluster memiliki jarak yang berdekatan antar objek.

2. Dasar Teori

2.1 Analisis RFM

Analisis RFM bertujuan untuk mengidentifikasi pelanggan yang memiliki loyalitas tinggi terhadap perusahaan berdasarkan proses transaksi. RFM dapat digunakan untuk melacak transaksi pelanggan terhadap waktu, frekuensi dan jumlah. RFM digunakan untuk menganalisis nilai pelanggan oleh perusahaan dengan tiga variabel yaitu R(*Recency*), F(*Frequency*), dan M(*Monetary*)[1]. Usaha untuk mendapatkan nilai

pelanggan secara khusus melalui tiga jenis indeks diatas, membantu perusahaan untuk mengetahui karakteristik pelanggan.

Proses segmentasi RFM menggunakan perhitungan bobot pada masing – masing variabel yaitu R, F, dan M. Penilaian bobot menggunakan angka lima sampai satu, dimana lima merupakan nilai tertinggi dan satu nilai terendah. Nilai tersebut dibagi rata sebesar 20% berdasarkan data yang akan diproses. Nilai *recency* tertinggi dilihat dari nilai rentang transaksi yang paling kecil. Nilai *frequency* dilihat dari yang paling banyak melakukan transaksi. Nilai *monetary* tertinggi dilihat dari jumlah dana transaksi terbesar.

Berikut definisi variabel RFM [2]:

1. Variabel R (*Recency*)

Recency merupakan rentang transaksi antara waktu pelanggan melakukan transaksi terakhir dengan waktu periode analisis yang dilakukan. Secara umum, semakin kecil nilai rentang transaksi pelanggan, pelanggan dianggap lebih mungkin untuk membeli sekali lagi.

2. Variabel F (*Frequency*)

Frequency merupakan jumlah pembelian yang dilakukan pelanggan selama waktu tertentu (dalam satu bulan, satu musim atau satu tahun). Semakin tinggi frekuensi ini, maka relatif lebih tinggi loyalitas pelanggan terhadap produk perusahaan ini. Nilai pelanggan di atas juga relatif penting untuk perusahaan. Oleh karena itu, nilai pelanggan untuk perusahaan lebih tinggi dari pada yang lain.

3. Variabel M (*Monetary*)

Monetary merupakan total biaya transaksi yang diterima perusahaan berdasarkan frekuensi pembelian yang dilakukan pelanggan selama periode tertentu. Secara umum, semakin banyak jumlah transaksi yang dilakukan pelanggan, semakin tinggi dan relatif penting nilai pelanggan tersebut untuk perusahaan. Sebaliknya, semakin sedikit jumlah transaksi yang dilakukan, semakin rendah juga nilai pelanggan tersebut terhadap perusahaan. Untuk *monetary*, nilai diperoleh dari perhitungan proses pembayaran yang telah dilakukan.

2.2 Clustering

Clustering atau Klasterisasi merupakan metode pengelompokan data berdasarkan kelas data yang telah ditentukan. Klasterisasi digolongkan sebagai metode *unsupervised learning*[3]. Prinsip dari klasterisasi adalah memaksimalkan kesamaan antar anggota dalam satu kelas dan meminimumkan kesamaan antar klaster.

Algoritma klasterisasi membutuhkan fungsi jarak untuk mengukur kemiripan antar data. Kelemahan pada metode ini adalah sering didapatkan hasil klaster yang tidak optimal karena pemilihan pusat klaster yang kurang baik dengan cara acak.

2.3 Clustering K-Means

Metode K-Means merupakan salah satu metode data klasterisasi untuk mempartisi data yang ada kedalam bentuk satu atau lebih klaster / kelompok[3]. Metode K-Means mempartisi data ke dalam klaster sehingga data yang memiliki karakteristik yang sama dikelompokkan kedalam satu klaster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Tujuan klasterisasi adalah untuk mengelompokkan objek hingga jarak tiap – tiap objek ke pusat kelompok di dalam suatu kelompok adalah minimum.

Proses pengelompokkan menggunakan K-Means secara umum dilakukan dengan algoritma dasar sebagai berikut:

1. Tentukan jumlah klaster.
2. Tentukan centroid awal dengan angka random.
3. Hitung jarak antar data yang diolah dengan angka random yang dipilih.
4. Alokasikan masing – masing data ke centroid/rata – rata terdekat. Lalu hitung nilai centroid baru dengan rumus sebagai berikut :

$$C = \frac{1}{M} \sum_{k=1}^M xk_i \quad (1)$$

Keterangan :

C = Centroid

M = Dimensi Obyek / Total Banyaknya Data

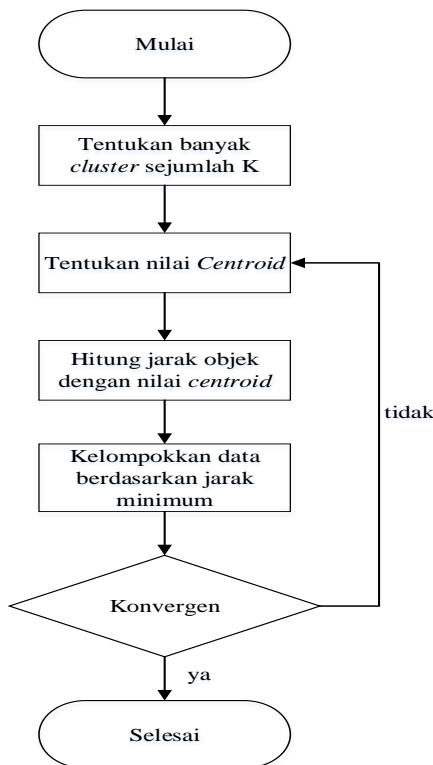
k = banyakobyek ($k=1,2,\dots,n$)

xk_i = feature vector ke- i ($i=1,2,\dots,n$)

5. Kembaliketahap 3, perhitunganakanselesaiikanilaiklaster yang lama danbarukonvergen. Dimanailaiklaster yang lama samadengannilaiklasterbaru.

Metode K-Means memiliki karakteristik sebagai berikut :

1. K-Means sangat cepat dalam proses klasterisasi dan sangat sensitive pada pembangkitan centroid awal secara acak.
2. Memungkinkan suatu klaster tidak memiliki anggota.
3. Hasil klasterisasi dengan K-Means bersifat unik / selaluberubah.



Gambar1 Flowchart Algoritma Clustering K-Means

2.4 Penentuan Centroid K-Means

Metode *K-Means* dalam penentuan *centroid* awal dilakukan secara acak tetapi hasil yang didapatkan kurang optimal, jika inisialisasi posisi awal sebagai pusat kluster didapat secara acak maka hasil kluster yang diperoleh tidak konsisten setiap melakukan perhitungan kluster. Pada saat melakukan perhitungan kluster dengan menentukan nilai *centroid* secara acak terkadang menghasilkan kluster yang sama namun hasilnya tetap tidak optimal. Oleh karena itu dalam inisialisasi *centroid* awal ditambahkan metode *Modified K-Means*.

Metode *Modified K-Means* merupakan hasil pengembangan dari algoritma *K-Means* dengan mengacu pada penelitian inisialisasi *centroid*. Cara kerja *Modified K-Means* dengan menyebarkan *centroid* awal di ruang fitur sehingga jarak antar *centroid* adalah sejauh mungkin[4]. Tujuannya agar hasil kluster yang terbentuk memiliki nilai *between* yang tinggi antar kluster. Algoritma *Modified K-Means* sebagai berikut :

1. Tentukan nilai M sebagai pusat data, misal $M(X_o, Y_o)$.

$$X_o = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

$$Y_o = \frac{1}{n} \sum_{i=1}^n y_i \quad (3)$$

Keterangan :

X_o = nilai sumbu X dari M.

Y_o = nilai sumbu Y dari M.

N = total data

2. Tentukan jumlah kluster, misal K.

3. Hitung jarak tiap data ke titik pusat M, misal M terletak pada data (X_o, Y_o) , maka jarak tiap piksel (X_p, Y_p) ke M (X_o, Y_o) .

$$D_m = \sqrt{(X_p - X_o)^2 + (Y_p - Y_o)^2} \quad (4)$$

Keterangan :

D_m = jarak data ke titik pusat M

X_p = nilai sumbu X pada data yang terukur

Y_p = nilai sumbu Y pada data yang terukur

4. Tentukan C1 sebagai pusat kluster pertama yang merupakan data terdekat dari pusat sebaran nilai M. jarak terpendek dari pusat M.

5. C2 ditentukan dengan melihat nilai sebaran data terdekat kedua dari M setelah C1. Lalu jarak antara C2-C1 harus lebih jauh dari jarak M-C1. bila tidak memenuhi maka ambil sebaran data terjauh dari M. Perhitungan menggunakan rumus Euclidean Distance (X_{p1}, Y_{p1})

$$D_m = \sqrt{(X_p - X_{p1})^2 + (Y_p - Y_{p1})^2} \quad (5)$$

Keterangan :

X_{p1} = nilai sumbu X pada data yang terukur

Y_{p1} = nilai sumbu Y pada data yang terukur

6. Lakukan langkah 4 dan 5 sampai k, dimana k = jumlah kluster untuk menentukan centroid.

7. Lakukan perhitungan untuk mencari nilai terdekat dengan masing – masing centroid lalu dikelompokkan.

8. Alokasikan masing – masing data ke centroid/rata – rata terdekat. Lalu hitung nilai centroid baru dengan rumus sebagai berikut:

$$C = \frac{1}{M} \sum_{k=1}^M x_k \quad (6)$$

C = Centroid

M = Dimensi Obyek / Total Banyaknya Data

k = banyak obyek (k=1,2,...n)

x_k = feature vector ke-i (i=1,2,...n)

9. perhitungan akan selesai jika nilai kluster yang lama sama dengan nilai kluster baru, bila belum sama

maka dilakukan pengulangan perhitungan pada tahap 7 dengan nilai centroid yang baru.

2.5 Euclidean Distance

Metode *Euclidean Distance* digunakan untuk menghitung jarak antar data dengan centroid [3]. Pengukuran ini didasarkan pada nilai obyek pada setiap k dimensi dalam pembelajaran. Euclidean Distance dapat menghitung jarak antar data sebanyak dua dimensi dan lebih. Rumus Euclidean Distance dengan 2 objek dan 2 dimensi:

$$D_{pq} = \sqrt{(p_2 - p_1)^2 + (q_2 - q_1)^2} \quad (7)$$

Keterangan :

D_{pq} = jarak 2 obyek

P₂ = nilai centroid pertama

P₁ = data pertama objek pertama

q₂ = nilai centroid kedua

q₁ = data pertama objek kedua

2.6 Evaluation Silhouette

Evaluation Silhouette merupakan salah satu metode untuk mengevaluasi suatu kluster. Metode ini digunakan untuk menguji kualitas kluster yang dihasilkan melalui proses klusterisasi [5]. Tahapan untuk menghitung nilai *Silhouette* adalah sebagai berikut :

1. Setiap obyek i, hitung rata – rata jarak dari obyek i dengan seluruh obyek yang berada dalam satu kluster. Hasil dari nilai rata – rata disebut a_i.
2. Setiap obyek i, hitung rata – rata jarak dari obyek i dengan obyek yang berada di kluster lainnya. Dari semua jarak rata – rata tersebut diambil nilai yang paling kecil. Nilai ini disebut b_i.
3. Setelah perhitungan pada tahap 1 dan 2 selesai maka didapatkan nilai silhouette coefficient:

$$S_i = (b_i - a_i) / \max(a_i, b_i) \quad (8)$$

Hasil perhitungan nilai silhouette coefficient dapat bervariasi antara -1 hingga 1. Hasil klusterisasi dikatakan baik jika nilai silhouette bernilai positif. Apabila nilai silhouette bernilai positif menandakan bahwa data yang terdapat pada kluster tersebut berada pada kelas atau kluster yang tepat.

2.7 Evaluation Variance

Evaluation Variance merupakan metode yang digunakan untuk mengevaluasi hasil kluster yang terbentuk [6]. Evaluasi Variance digunakan untuk melihat jarak nilai data dalam satu kluster [6]. Jarak data yang kecil atau mendekati 0, menunjukkan bahwa data dalam satu kluster memiliki kemiripan nilai yang dekat antar data. Rumus Varians sebagai berikut :

$$R = \frac{1/C \sum_{i=1}^c R_c}{1/C \sum_{i,j=1}^c R_{ij}} \quad (9)$$

Keterangan :

C = total jumlah kluster

i = data ke-i dalam kluster

R_c = jarak rata-rata objek dalam satu kluster

R_{ij} = jarak rata-rata objek kluster dengan kluster lain

3. Hasil Percobaan

Pengujian program prediksi *customer* menggunakan sistem segmentasi RFM dan *clustering K-Means* terdiri dari dua tahap pengujian. Dua tahap pengujian tersebut, yaitu pengujian terhadap modul program dan pengujian terhadap kinerja rancangan program.

Pengujian terhadap modul bertujuan untuk menguji apakah setiap modul dalam aplikasi berjalan dengan baik. Hasil pengujian terhadap modul menunjukkan bahwa semua modul berjalan dengan baik dan sesuai rancangan. Berikut merupakan modul dalam aplikasi :

1. Modul Main Menu
2. Modul Pelanggan
3. Modul Transaksi Pelanggan
4. Modul Segmentasi

Pengujian terhadap kinerja program dilakukan dengan membandingkan hasil perhitungan manual dan perhitungan program. Perhitungan dilakukan dengan menggunakan data periode Januari 2013. Hasil perhitungan yang dilakukan menghasilkan hasil yang sama baik dari perhitungan manual dan perhitungan program. Perhitungan dilakukan pada proses input dua dan tiga kluster. Hasil dilihat dari nilai evaluasi *silhouette* pada tabel 1.

Tabel 1 Evaluasi *Global Silhouette*

Cluster	Global Silhouette (program)	Global Silhouette (manual)
2	0,1211	0,1211
3	0,3529	0,3529

Pengujian berikutnya melakukan pengujian pada data perusahaan periode Januari 2013 – Desember 2014. Pengujian dilakukan dengan melakukan empat evaluasi, yaitu **Evaluasi *Global Silhouette*, Evaluasi *Variance*, *LineChart*, *Scatter Plot*, dan Evaluasi terhadap data asli.**

3.1 Evaluasi *Global Silhouette*

Pengujian dilakukan terhadap data periode Januari 2013 Desember 2014. Pengujian pertama yang dilakukan adalah pengujian terhadap nilai evaluasi *Global Silhouette*. Pengujian dilakukan terhadap dua kluster dan tiga kluster. Hasil yang baik dan tepat adalah hasil yang memiliki nilai *Global Silhouette* bernilai positif. Hasil yang positif menunjukkan bahwa data yang terdapat

pada kluster tersebut berada pada kluster yang benar. Hasil positif juga menunjukkan bahwa jarak rata – rata antar objek dalam satu kluster lebih kecil dari jarak rata – rata antar objek dengan objek yang berada pada kluster lain.

i. Dua Kluster

Pengujian pertama dilakukan dengan input dua kluster. Nilai *Silhouette* pada pembentukan dua kluster adalah **0,2839** menunjukkan hasil kluster yang baik. Hasil bernilai positif menunjukkan bahwa hasil klusterisasi mengeluarkan hasil pengelompokan yang baik. Hasil baik yang dimaksud adalah anggota yang terdapat pada kluster satu dan kluster dua berada dikelas yang tepat.

ii. Tiga Kluster

Pengujian kedua dilakukan dengan input tiga kluster. Nilai *Global Silhouette* pada input tiga kluster sebesar **0,1157**. Nilai tersebut menunjukkan nilai yang lebih rendah pada saat proses input dua kluster, namun hasil tetap bernilai positif. Sehingga proses klusterisasi bekerja dengan baik. Baik yang dimaksud adalah jarak rata – rata antar objek dalam satu kluster lebih kecil dari jarak rata – rata antar objek dengan objek yang berada pada kluster lain dan anggota pada masing - masing kelas berada di kelompok atau cluster yang benar. Berikut merupakan tabel hasil nilai evaluasi *silhouette* dari kedua pengujian tersebut dapat dilihat pada Tabel 2.

Tabel 2 Nilai Evaluasi *Global Silhouette*

nilai <i>Global Silhouette</i>	
dua kluster	tiga kluster
0,2839	0,1157

3.2 Evaluasi *Variance*

Pengujian evaluasi varians digunakan untuk mengecek kualitas kluster yang terbentuk. Kualitas kluster yang baik adalah kluster yang memiliki nilai jarak terkecil dalam kluster. Nilai jarak yang kecil pada proses perhitungan kluster menunjukkan bahwa kemiripan nilai data dalam suatu kluster memiliki kemiripan nilai yang dekat. Hasil perhitungan varians yang menunjukkan kluster yang paling baik adalah nilai yang mendekati nilai 0.

i. Dua Kluster

Pengujian pertama dilakukan pada proses pembentukan dua kluster. Hasil nilai evaluasi varians pada pembentukan dua kluster adalah **2,1583**. Hasil tersebut menunjukkan nilai yang cukup baik karena angka tidak jauh dari angka 0. Hasil varians pada pembentukan dua kluster menunjukkan bahwa data yang terdapat pada kluster tersebut memiliki keberagaman nilai yang mirip.

ii. Tiga Kluster

Pengujian kedua dilakukan pada proses pembentukan tiga kluster. Hasil nilai evaluasi varians pada pembentukan tiga kluster adalah **2,8638**. Nilai varians pada pembentukan tiga kluster menghasilkan nilai yang lebih besar dari pembentukan dua kluster. Nilai varians yang lebih besar menunjukkan bahwa data yang terdapat pada proses pembentukan tiga kluster memiliki keberagaman nilai yang cukup berbeda dibandingkan pada pembentukan dua kluster.

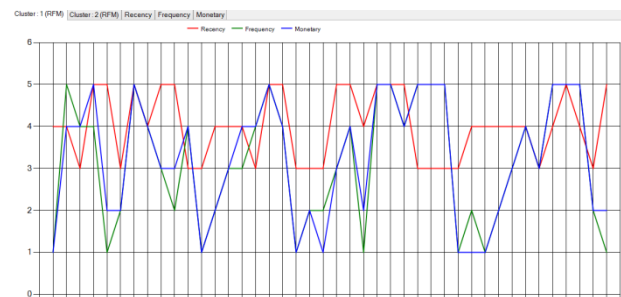
Berikut merupakan tabel hasil nilai evaluasi *silhouette* dari kedua pengujian tersebut dapat dilihat pada Tabel 3.

Tabel 3 Nilai Evaluasi *Variance*

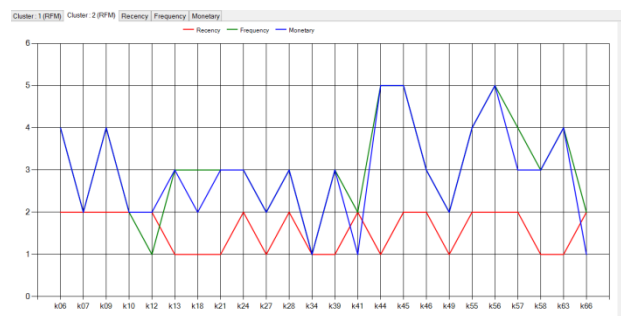
Nilai <i>Evaluation Variance</i>	
dua kluster	tiga kluster
2,1583	2,8638

3.3 Evaluasi *Graph Chart* (dua kluster dan tiga kluster)

Hasil gambar *chart* pada proses pembentukan dua kluster dapat dilihat pada Gambar 2 dan 3.

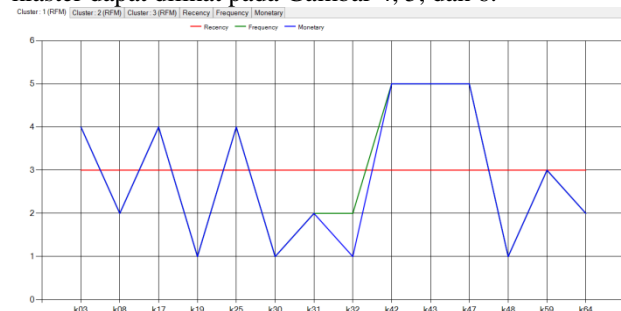


Gambar 2 Cluster : 1 (RFM)

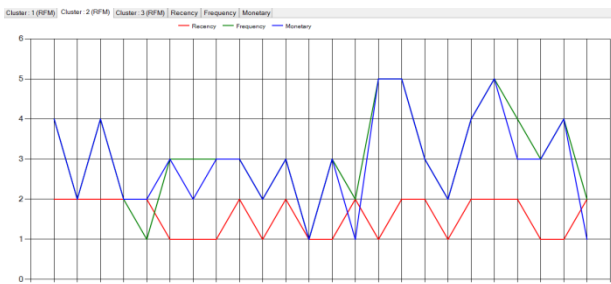


Gambar 3 Cluster : 2 (RFM)

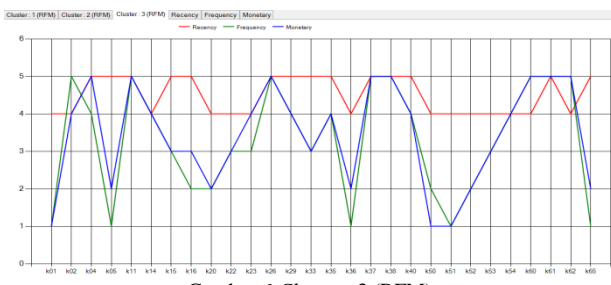
Hasil gambar *chart* pada proses pembentukan tiga kluster dapat dilihat pada Gambar 4, 5, dan 6.



Gambar 4 Cluster : 1 (RFM)



Gambar 5 Cluster : 2 (RFM)



Gambar 6 Cluster : 3 (RFM)

Kesimpulan dari *output* gambar *chart* pada proses pembentukandua kluster dan tiga kluster adalah hasil terbaik didapat pada saat pembentukantiga kluster. Hasil pada kluster ketiga menunjukkan nilai RFM yang tinggi, hal ini menunjukkan bahwa pelanggan pada kluster tersebut memiliki tingkat loyalitas yang tinggi terhadap perusahaan dan dapat dijadikan sebagai bahan pertimbangan bagi perusahaan untuk memberikan diskon terhadap pelanggan yang terdapat pada kluster tersebut.

3.4 Evaluasi Scatter Plot

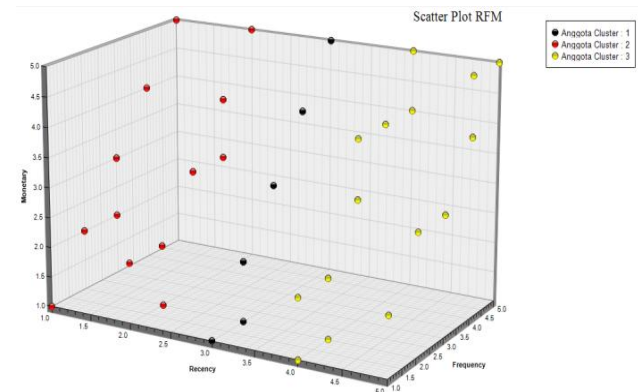
Pengujian ketiga terhadap data periode Januari 2013 Desember 2014 dilakukan dengan melihat gambar *scatter plot* terhadap dua kluster dan tiga kluster. Hasil gambar yang baik dilihat dengan cara warna pada masing – masing kluster mengelompok berdekatan atau tidak. Apabila warna antar kluster terkelompok secara berdekatan dapat disimpulkan bahwa pelanggan yang terkelompok pada kluster memiliki kemiripan nilai RFM yang berdekatan.

Hasil gambar *scatter plot* pada proses *input* dua kluster dapat dilihat pada Gambar 7.

Gambar 7 Scatter Plot 3D (input dua kluster)

Pada Gambar 7 menunjukkan titik warna merah merupakan pelanggan yang terdapat pada kluster satu dan warna hitam merupakan pelanggan pada kluster dua. Hasil gambar tersebut dapat dilihat bahwa hasil kluster menunjukkan hasil yang cukup baik. Seluruh pelanggan mengelompok cukup berdekatan. Namun anggota pada kluster satu terjadi penyebaran pelanggan terjadi cukup jauh. Sehingga nilai RFM pada kluster satu memiliki keberagaman nilai.

Hasil *scatter plot* pada proses *input* tiga kluster dapat dilihat pada Gambar 8.



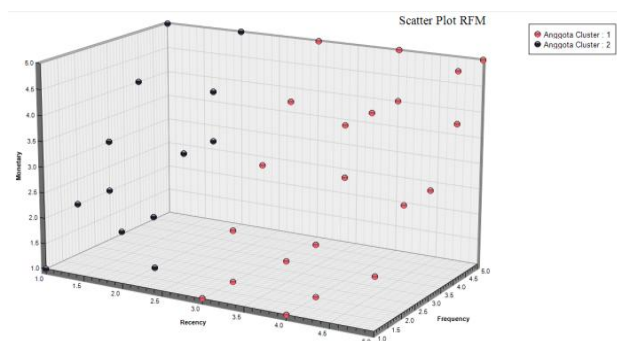
Gambar 8 Scatter Plot 3D (input tiga kluster)

Pada Gambar 8 menunjukkan hasil dari pembentukan tiga kluster. Pada gambar ini dapat dilihat bahwa hasil klusterisasi bekerja dengan baik. Hasil *scatter plot* pada ketiga warna tersebut memperlihatkan bahwa pelanggan yang memiliki kemiripan nilai RFM terkelompok dengan baik. Jarak antar warna yang cukup jauh menandakan bahwa masing – masing kluster memiliki keberagaman nilai RFM yang berbeda.

Dari hasil pembahasan yang dilakukan pada proses *input* dua kluster dan tiga kluster menunjukkan bahwa hasil pengelompokkan terbaik terdapat pada proses pembentukan tiga kluster. Hasil tersebut menunjukkan nilai RFM pada satu kluster memiliki keberagaman nilai yang dekat dan nilai antar kluster memiliki keberagaman nilai RFM yang berbeda. Sehingga hasil *output* pada kluster ketiga memnunjukkan hasil pengelompokkan yang baik.

3.5 Evaluasi dengan Data Asli

Tahap pengujian terakhir adalah dengan melakukan perbandingan hasil pembentukan dua kluster dan tiga kluster dengan data pelanggan yang diberikan diskon oleh perusahaan. Tujuan dari pengujian tersebut untuk melihat apakah data pelanggan yang diberikan diskon oleh perusahaan pada periode Januari 2013 – Desember 2014 memiliki kesamaan dengan hasil kluster pada program. Hasil kluster yang memiliki kemiripan yang dekat dengan data pelanggan yang sudah diberikan diskon akan dijadikan sebagai bahan masukan kepada perusahaan untuk pemberian diskon. Berikut hasil



perbandingan dengan data asli perusahaan dapat dilihat pada Tabel 2.

Tabel 2 Perbandingan Data Perusahaan dan Data Kluster

Input Program Dua Kluster					
Data Pelanggan Perusahaan periode Januari 2013 - Desember 2014	Data Diskon Perusahaan	Kluster Ke :	Data Hasil Kluster	Data Diskon	Tingkat Kemiripan Dengan Data Asli(%)
66	25	1	43	14	56%
		2	23	11	44%
Input Program Tiga Kluster					
Data Pelanggan Perusahaan periode Januari 2013 - Desember 2014	Data Diskon Perusahaan	Kluster Ke :	Data Hasil Kluster	Data Diskon	Tingkat Kemiripan Dengan Data Asli(%)
66	25	1	14	5	20%
		2	24	4	16%
		3	28	16	64%

Keterangan :

Data pelanggan perusahaan : banyaknya pelanggan yang melakukan transaksi periode Januari 2013 – Desember 2014

Data diskon perusahaan : jumlah banyak pelanggan yang diberikan diskon oleh perusahaan sebanyak 25 pelanggan periode Januari 2013 – Desember 2014

Kluster ke : jumlah *input cluster* dari program segmentasi pelanggan

Data hasil kluster : banyak data pelanggan yang terdapat pada kluster/kelompok tersebut.

Data diskon : banyaknya data pelanggan yang terdapat pada data diskon perusahaan periode Januari 2013 – Desember 2014.

Presentasi kemiripan data : presentase seberapa besar data pelanggan yang terdapat pada data diskon pelanggan yang diberikan oleh perusahaan.

Hasil dari tabel tersebut dapat disimpulkan bahwa hasil kluster yang paling sama dengan data diskon yang diberikan oleh perusahaan selama periode dua tahun adalah pada saat *input* tiga kluster dan terdapat pada kluster tiga. Hal ini dapat dijadikan sebagai bahan masukan dan pertimbangan kepada perusahaan untuk memberikan diskon terhadap seluruh pelanggan yang terdapat pada kluster tersebut.

5. Kesimpulan

Kesimpulan yang dapat diperoleh dari hasil pengujian yang dilakukan terhadap system prediksi loyalitas pelanggan dengan metode segmentasi RFM dan pengelompokan dengan algoritma *Clustering Modified K-Means* adalah sebagai berikut:

1. Aplikasi segmentasi pelanggan telah berfungsi dengan baik setiap tombolnya dan sesuai dengan spesifikasi rancangan yang telah ditentukan.
2. Metode segmentasi RFM dan *clustering modified k-means* dapat diimplementasikan pada kasus perusahaan.
3. Hasil *clustering* pembentukan dua kluster menunjukkan ciri pelanggan yang cukup sering belanja sampai periode Desember 2014 dengan

perusahaan dan pembentukan tiga kluster menghasilkan ciri pelanggan yang berbelanja dalam nominal yang cukup besar dengan frekuensi yang cukup besar.

4. Hasil pengujian evaluasi *silhouette* dan *variance* menunjukkan hasil terbaik pada proses pembentukan dua kluster. nilai *silhouette* sebesar **0,2839** dan nilai *variance* sebesar **2,1583**.
5. Hasil pengujian berdasarkan gambar *LineChart* dan *ScatterPlot* menunjukkan hasil terbaik pada proses pembentukan tiga kluster.
6. Hasil *clustering* yang menghasilkan data yang mirip pada data diskon perusahaan periode Januari 2013 – Desember 2014 adalah pada proses pembentukan tiga kluster dengan tingkat kemiripan sebesar 64%.
7. *Clustering K-Means* membantu perusahaan dalam mencari pelanggan yang memiliki nilai loyalitas yang tinggi terhadap perusahaan.

REFERENSI

- [1] Kohavi. R; Parekh. R, "Visualizing RFM Segmentation", USA, 2004, diakses :10 2 2015.
- [2] Cheng C. Hsue; C. Y. Shyang, "Jurnal Expert System with Application: An International Journal," *Classifying the segmentation of customer value via RFM Model*, vol. 36, pp. 4176-4184, 2009.
- [3] Mirkin, B., "*clustering for data mining : A data recovery approach*", (London: Chapman & Hall/CRC, 2005)
- [4] Barakbah, Ali R.; Helen, Afrida. "Seminar Nasional Soft Computing Intelligent System and Information," *Optimized K-Means: An Algorithm of Initial Centroid Optimization for K-Means*, 2005.
- [5] Peter. J. Rousseeuw, *Silhouette: a Graphical Aid to the interpretation and Validation of Cluster Analysis*, 1987
- [6] Handhayani, T; et.al, *Kernel Based Integration of Gene expression and DNA copy number*, publisher: IEEE, 2013

Herwin, merupakan mahasiswa program sarjana S1, program studi Teknik Informatika, Fakultas Teknologi Informasi Universitas Tarumanagara.