

Penerapan Data Mining Menggunakan Metode K-Means Clustering Untuk Pengelompokan Data Pelanggan (Studi Kasus : PT. Pinus Merah Abadi)

Hesti Astuti

Sistem Informasi, STMIK WIT

Abstract

This data mining research uses K-Means Clustering method to classify customer data that produces the number of clusters and customer characteristic from each cluster. As well as using RFM analysis (Recency, Frequency, Monetary) which presents the number of customers in each cluster. Data used in the process of classifying customers is the sales transaction data from January to September 2017. The result of this research is to produce three customer cluster with customer characteristic from each cluster. Cluster 1 amounted to 679 with low recency ↓ low frequency ↓ low monetary ↓ included into new customers, cluster 2 amounted to 17 with high recency ↑ high frequency ↑ high monetary ↑ included into loyal customers and cluster 3 amounted to 334 with high recency ↑ low frequency ↓ low monetary ↓ including into potentially inactive loyal customers. The results of this clustering can help PT. Eternal Red Pinus to adopt policies and strategies to improve or retain customers in accordance with the customer cluster.

Keywords : Customer, Data mining, K-Means, Clustering, RFM (Recency, Frequency, Monetary)

Abstrak

Penelitian data mining ini menggunakan metode *K-Means Clustering* untuk mengelompokkan data pelanggan yang menghasilkan jumlah *cluster* dan karakteristik pelanggan dari masing-masing *cluster*. Serta menggunakan analisis RFM (*Recency, Frequency, Monetary*) yang mempresentasikan jumlah pelanggan disetiap *cluster*. Data yang digunakan dalam proses pengelompokan pelanggan adalah data transaksi penjualan Januari sampai September 2017. Hasil dari penelitian ini adalah menghasilkan tiga *cluster* pelanggan dengan karakteristik pelanggan dari tiap-tiap *cluster*. *Cluster 1* berjumlah 679 dengan kriteria *recency* rendah ↓ *frequency* rendah ↓ *monetary* rendah ↓ termasuk kedalam pelanggan baru, *cluster 2* berjumlah 17 dengan kriteria *recency* tinggi ↑ *frequency* tinggi ↑ *monetary* tinggi ↑ termasuk ke dalam pelanggan yang loyal dan *cluster 3* berjumlah 334 dengan kriteria *recency* tinggi ↑ *frequency* rendah ↓ *monetary* rendah ↓ termasuk kedalam pelanggan tidak loyal yang berpotensi tidak aktif. Hasil dari *clustering* ini dapat membantu PT. Pinus Merah Abadi untuk mengambil kebijakan dan strategi dalam meningkatkan atau mempertahankan pelanggan yang sesuai dengan *cluster* pelanggan tersebut.

Kata Kunci : Pelanggan, Data mining, K-Means, Clustering, RFM (*Recency, Frequency, Monetary*)

1. Pendahuluan

Data mining ini dapat dipilih sebagai sistem yang tepat untuk kemajuan perusahaan yang dapat menemukan informasi yang berada di gudang data perusahaan yang terletak pada penyimpanan data komputer dan data bisnis tersebut dapat bermanfaat sebagai informasi dalam setiap langkah pembangunan. Dan perlu diketahui data mining merupakan suatu alat yang memungkinkan para pengguna untuk mengakses secara data dengan jumlah yang sangat besar.

K-Means merupakan salah satu metode data non-hierarchical clustering yang dapat mengelompokkan data ke dalam beberapa cluster berdasarkan kemiripan dari data tersebut sehingga data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster dan yang memiliki karakteristik yang berbeda dikelompokkan dalam cluster yang lain yang memiliki karakteristik yang sama.

Dalam sebuah usaha penjualan dan distributor persaingan merupakan hal yang wajar dan tidak bisa dihindari terutama antar perusahaan yang mempunyai segmentasi pasar yang sejenis. Ketatnya persaingan antar

perusahaan ini menuntut perusahaan untuk berusaha lebih maju dan terbaik dalam memberikan pelayanan yang membuat pelanggan puas agar tidak beralih ke perusahaan lainnya.

Pengelompokkan pelanggan berdasarkan karakteristik masing-masing akan berpengaruh pada manajemen pemasaran perusahaan. Dengan adanya pengelompokan ini akan diketahui data pelanggan dengan karakteristik yang sama dikelompokkan kedalam satu cluster dan yang berbeda dikelompokkan dalam cluster yang lain.

Data pelanggan cenderung merupakan data transaksi pelanggan yang memiliki banyak atribut. Untuk mengatasi beberapa kekurangan dalam pemilihan atribut yang ada, diusulkan penggunaan metode RFM (*Recency, Frequency, Monetary*).

Oleh karena itu diperlukan teknik data mining untuk pengelompokkan data pelanggan pada PT. Pinus Merah Abadi yaitu menggunakan K-Means Clustering, sehingga PT. Pinus Merah Abadi dapat mengambil kebijakan dan strategi dalam meningkatkan kualitas pelayanan dan dan

dalam mempertahankan pelanggan serta dapat reward yang sesuai dengan jenis cluster pelanggan tersebut.

1.1. Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah :

1. Berapa *cluster* dan bagaimanakah karakteristik dari setiap *cluster* yang terbentuk?
2. Variabel manakah yang paling dominan yang menjadi pembeda antar *cluster*?

1.2. Batasan Masalah

1. Data yang digunakan data transaksi penjualan yang diambil dari bulan Januari sampai September tahun 2017
2. Atribut yang digunakan adalah menggunakan metode analisis RFM (*recency, frequency, dan monetary*)

1.3. Tujuan

Tujuan yang ingin dicapai dari penelitian ini adalah menerapkan data mining menggunakan algoritma k-means clustering untuk mengelompokkan data pelanggan pada PT. Pinus Merah Abadi menjadi beberapa cluster yang sesuai dengan karakteristik yang dibentuk dari masing-masing pelanggan.

2. Kerangka Teori

2.1. Data Mining

Data mining memiliki beberapa istilah padanan, yaitu knowledge discovery ataupun pattern recognition. Kedua istilah ini memiliki ketepatan masing-masing. "Istilah knowledge discovery atau penemuan pengetahuan tepat digunakan karena tujuan utama dari data mining memang untuk mendapatkan pengetahuan yang masih tersembunyi didalam bongkahan data." (Susanto & Suryadi, 2010:2)

Salah satu teknik yang dibuat data mining adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data lain yang tidak berada dalam basis data yang tersimpan. Dalam data mining, pengelompokan data juga bisa dilakukan, kebutuhan untuk prediksi juga dapat memanfaatkan teknik ini. Tujuannya adalah agar dapat mengetahui pola universal data-data yang ada.

2.2. Knowledge Discovery In Database (KDD)

Knowledge Discovery In Database (KDD) adalah metode untuk memperoleh untuk memperoleh pengetahuan dari database yang terdapat tabel-tabel yang saling berelasi. Knowledge Discovery In Database (KDD) berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah kumpulan data. Hasil yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan (knowledge base) untuk keperluan pengambilan keputusan.

Proses Knowledge Discovery In Database (KDD) secara garis besar dapat dijelaskan sebagai berikut : (Kusrini & Emha :2009)

1. Data Selection

Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *Knowledge Discovery In Database (KDD)* dimulai.

2. Pre-processing/Cleaning

Pemrosesan pendahuluan dan pembersihan data merupakan operasi dasar seperti penghapusan noise dilakukan. Sebelum proses data mining dapat dilaksanakan, perlu dilakukannya proses cleaning pada data yang menjadi fokus *Knowledge Discovery In Database (KDD)*. Proses ini mencakup membuang duplikasi data, memeriksa data inkonsisten, dan memperbaiki kesalahan pada data.

3. Transformation

Tahap ini yaitu pencarian fitur-fitur yang berguna untuk mempresentasikan data bergantung kepada goal yang ingin dicapai.

4. Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi pemilihan metode atau algoritma sangat bergantung pada tujuan dan proses *Knowledge Discovery In Database (KDD)*.

5. Interpretation/Evaluation

Tahap ini penerjemahan pola-pola yang dihasilkan dari data mining. Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses *Knowledge Discovery in Database (KDD)* yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

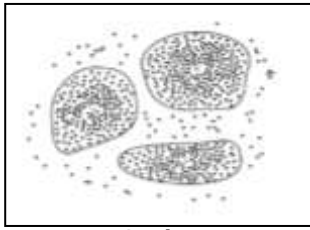
2.3. Clustering

Pada dasarnya *clustering* merupakan suatu metode untuk mencari atau mengelompokkan data yang memiliki kemiripan karakteristik antara satu data dengan data yang lain. *Clustering* adalah proses pengelompokan item data ke dalam sejumlah kecil grup sedemikian sehingga masing-masing grup mempunyai sesuatu persamaan yang esensial.

Clustering dikategorikan kedalam teknik "*Undirect Knowledge*" atau "*Unsupervised Learning*" karena tidak membutuhkan proses pelatihan untuk klasifikasi awal data dalam masing-masing kelompok atau *cluster*.

Pengelompokkan mengacu pada pengelompokan kumpulan catatan, atau kasus ke dalam objek yang serupa. "*Cluster* adalah kumpulan catatan yang mirip satu sama lain dan tidak sesuai dengan catatan dalam kelompok lain. (Daniel T. Larose, 2005: 147)

"Beberapa algoritma pada teknik clustering menggunakan perhitungan jarak minimum untuk mengukur kemiripan antar data apabila data berupa data numeric." (Kusnawi, 2007: 6).



Gambar 2.1

Contoh Klasterisasi (Kusnawi, 2007: 6)

2.4. K-Means

Algoritma *K-Means* merupakan implementasi algoritma *clustering* partisional yang paling sederhana dan paling banyak digunakan. *K-Means* juga menggunakan kuadrat *error criterion*. Algoritma ini mulai mempartisi ruang data secara acak sambil menunjukan sampel yang adad kedalam kluster-kluster berdasarkan kemiripan antara kluster dan sampel, sampai sebuah *criterion* yang *convergen* ditemukan.

Algoritma *K-Means* merupakan proses secara berulang-ulang untuk mendapatkan basis atau *cluster*. “*Metode K-Means merupakan algoritma klasterisasi yang paling tua dan paling banyak digunakan dalam berbagai aplikasi kecil hingga menengah karena kemudahan implementasinya*”. (Suyanto, 2017: 262)

Menurut Santosa (2007), langkah-langkah melakukan clustering dengan metode k-means adalah sebagai berikut:

1. Pilih jumlah *cluster* *k*.
2. Inisialisasi *k* pusat *cluster* ini bisa dilakukan beberapa cara. Namun paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka random.
3. Alokasikan semua data atau objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan semua data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Jarak antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

$$D(i,j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (1)$$

Dimana :

$D(i,j)$ = jarak data ke *i* ke pusat *cluster* *j*

X_{ki} = data ke *i* pada atribut data ke *k*

X_{kj} = titik pusat ke *j* pada atribut ke *k*

4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data atau objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari *cluster* tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
5. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali lagi ke langkah nomor tiga sampai pusat *cluster* tidak berubah lagi.

2.5. Analisis RFM (*Recency, Frequency, Monetary*)

Model analisa RFM dikembangkan pertama kali oleh Hughes sebagai model untuk menganalisa nilai pelanggan. Model ini membedakan pelanggan yang penting dari sejumlah data dengan menggunakan tiga atribut yaitu selang waktu (interval) pemakaian pelanggan, frekuensi dan jumlah uang. Menggunakan informasi tentang perilaku oelanggan masa lalu yang mudah dilacak dan tersedia.

Tujuan RFM adalah untuk meramalkan perilaku konsumen di masa yang depan (mengarahkan keputusan yang lebih baik). Oleh karena itu, perlu menterjemahkan perilaku pelanggan dalam “angka” sehingga dapat digunakan sepanjang waktu.

Menurut jurnal Herwin, Teny Handhayani, Dyah E Herwindiati. 2015. Dikatakan bahwa analisis RFM bertujuan untuk mengidentifikasi pelanggan yang memiliki loyalitas tinggi terhadap berdasarkan proses transaksi.

1. *Recency*, yaitu rentang waktu (dalam satuan hari, bulan, tahun) daritransaksi terakhir yang dilakukan oleh konsumen sampai saat ini.
2. *Frequency*, yaitu jumlah total transaksi atau jumlah rata-rata transaksi dalam satu periode.
3. *Monetary*, yaitu jumlah rata-rata nilaipembelian konsumen dalam suatu satuan waktu

2.6. SPSS Statistic

SPSS merupakan salah satu software statistika yang paling populer digunakan di Indonesia. User atau para pengguna lebih memilih SPSS karena selain bahasa SPSS yang cepat dan mudah untuk dipahami dan tampilan outputnya terkesan lebih bagus.

Sebenarnya nama SPSS sudah berubah dari tahun 2009 menjadi PASW, namun nama SPSS lebih dikenal oleh kalangan umum. “SPSS atau kependekan dari Statistical Package for the Social Sciences adalah paket software untuk analisis statistika dan manajemen”. (Getut Prameswari, 2014:9). SPSS mengolah data dalam menyelesaikan permasalahan statistika seperti bidang di bisnis dan penelitian menjadikan SPSS menjadi alat analitik dan prediktif.

2.7. Proses Standarisasi Data

Proses Standarisasi data dilakukan karena data sangat bervariasi dalam satuan. Perbedaan yang sangat mencolok akan menyebabkan bias dalam menganalisis faktor sehingga data asli harus di standarisasi sebelum dianalisis. Proses Standarisasi data bisa dilakukan dengan data diubah dalam bentuk Z-Score. “Z-Score merupakan unit-unit simpangan baku (standard deviation) untuk mengukur perbedaan tersebut”. (Singgih Susanto:2015)

Untuk mengetahui nilai zscore didapat dengan menggunakan rumus:

$$z_i = \frac{x_i - \bar{x}}{s} \quad (2)$$

Dimana:

i = data ke 1, 2, 3, ..., *n*.

Xi= nilai rata-rata

S= nilai standarisasi

Nilai yang terkait dengan proses standarisasi data yang mengacu pada angka z mempunyai dua bagian yaitu tanda (bisa positif atau negatif) dan nilai numerik. Kondisi diatas rata-rata total diberi tanda positif dan dibawah rata-rata total diberi tanda negatif. Untuk mengetahui tafsiran nilai atau angka rata-rata pada variabel Z-Score dapat digunakan rumus:

$$x = \mu + z \cdot \sigma \quad (3)$$

Dimana:

X= rata-rata sampel

μ = standar deviasi

Z= nilai standarisasi yang didapat pada SPSS

3. Metodologi

Kerangka kerja penelitian ini dapat dilihat pada gambar 3.1 berikut ini :



Gambar 3.1 Metode Penelitian

Langkah-langkah penelitian dapat dijelaskan sebagai berikut:

1. Pengumpulan Data

Pengumpulan data penelitian ini dilakukan dengan cara sebagai berikut:

a. Observasi

Teknik pengumpulan data dengan cara mengamati objek secara langsung pada objek yang sedang diteliti.

b. Wawancara

Teknik pengumpulan data dengan cara berkomunikasi secara langsung atau tanya jawab kepada pegawai PT. Pinus Merah Abadi.

2. Studi Literature

Metode ini membantu mencari teori dasar yang diperlukan dalam penelitian, seperti tentang data mining dan metode *K-Means Clustering*, pengumpulan data training dan data testing untuk proses pengeompokan pelanggan.

3. Seleksi Data

Pada tahapan ini dilakukan untuk langkah awal dalam melakukan proses data mining. Tahap ini dilakukan untuk mendapatkan kolom-kolom data yang tepat yang disebut atribut. Data yang digunakan dalam penelitian ini berasal dari data transaksi penjualan bulan Januari-September tahun 2017. Format data yang digunakan adalah format .xlsx (*Excel Document*).

Tabel 3.1 Tabel Transaksi Januari-September 2017

No.	Nama Atribut	Keterangan
1	Outlet	Nomor Outlet Pelanggan
2	Nama Outlet	Nama Outlet Pelanggan
3	Slsno	Nomor Sales
4	Tgl Faktur	Tanggal Faktur Transaksi Pelanggan
5	Jumlah Bruto	Jumlah Biaya Transaksi
6	Discount	Diskon yang diperoleh
7	PPN	Pemotongan Jumlah PPN yang dibayar
8	Total	Total Biaya Transaksi

Tabel 3.2 Cuplikan Tabel Transaksi Januari-September 2017

NO.	OUTLET	NAMA OUTLET	SLSNO	TGLFAKTUR	NOFAKTUR	JMLBRUTTO	DISCOUNT	PPN	TOTAL
1	641116	REZA TK	648602	12-01-2017	16326215	Rp. 79.444	Rp. -	Rp. 7.944	Rp. 87.388
2	641120	AGUS TK	648602	12-01-2017	16326216	Rp. 330.850	Rp. 4.963	Rp. 32.583	Rp. 368.315
3	641125	APRI TK	648602	12-01-2017	16326217	Rp. 83.354	Rp. -	Rp. 8.335	Rp. 102.689

4. Pre-processing/Cleaning Data

Tahap ini akan meliputi tahapan membuang duplikasi data, membuang data yang tidak konsisten atau noise, duplikasi data, memperbaiki kesalahan data. Setelah dilakukan seleksi data diatas, yaitu dengan mengurangi atribut yang tidak diperlukan untuk memudahkan pada saat melakukan tahap data mining selanjutnya. Pemilihan atribut yang dibutuhkan dalam proses clustering pelanggan menggunakan metode RFM yaitu berkaitan dengan *recency*, *frequency*, dan *monetary*.

Tabel 3.3 Pemilihan Atribut dari Tabel Transaksi

Field	Keterangan
Outelt	Nomor Outlet
Nama Outlet	Nama Outlet Pelanggan
Tanggal Faktur	Menandakan <i>recency</i> , tanggal transaksi terakhir yang dilakukan oleh pelanggan
Frekuensi Beli	Menandakan <i>frequency</i> , jumlah transaksi selama periode yang ditentukan
Total	Menandakan <i>monetary</i> , jumlah uang selama periode yang ditentukan.

a. Kriteria R (*recency*)

Membutuhkan atribut yang menunjukkan adanya rentang waktu transaksi terakhir pelanggan dengan periode analisis, sehingga atribut yang dibutuhkan adalah atribut tanggal transaksi.



Gambar 3.2 Kriteria *recency*

b. Kriteria F (*frequency*)

Atribut yang dibutuhkan adalah atribut tanggal transaksi yang dihitung jumlahnya.



Gambar 3.3 Kriteria *frequency*

c. Kriteria M (monetary)

Membutuhkan atribut yang berhubungan dengan biaya yang telah dihabiskan pelanggan selama melakukan transaksi, atribut yang dibutuhkan adalah atribut Total yang dibayarkan pelanggan.



Gambar 3.4 Kriteria monetary

Tabel 3.4 Cuplikan Tabel Setelah Pra-proses

NOOUTLET	NAMAOUTLET	TGL TERAKHIR	JUML TRANSAKSI	TOTAL
640000	PENJUALAN KANTOR	12/09/2017	3	Rp. 675.123
640001	PENJUALAN TOPPING U	24/05/2017	3	Rp. 977.719
640002	PENJUALAN NED 1+1	12/09/2017	3	Rp. 640.921

5. Transformasi Data

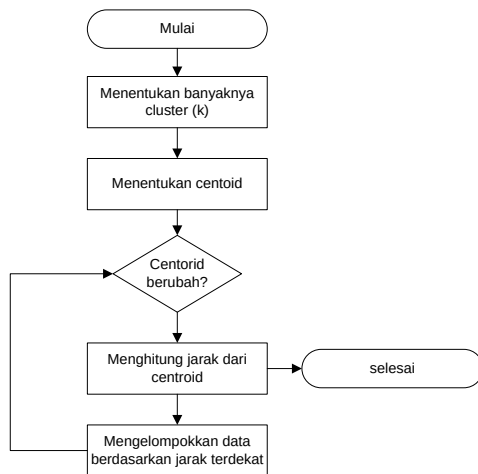
Pada tahap ini, akan dilakukan proses pengubahan data ke dalam format yang sesuai untuk diproses dalam data mining.

Tabel 3.5 Cuplikan Tabel RFM telah ditransformasi

NO	NOOUTLET	NAMAOUTLET	RECENCY	FREQUENCY	MONETARY
1	640000	PENJUALAN KANTOR	18	3	Rp. 675.123
2	640001	PENJUALAN TOPPING U	129	3	Rp. 977.719
3	640002	PENJUALAN NED 1+1	18	3	Rp. 640.921

6. Penerapan K-means Clustering

Pada umumnya kinerja metode k-means secara berurutan adalah sebagai berikut:



Gambar 3.5. Alur Algoritma K-Means

7. Menguji Hasil

Tahap ini adalah suatu teknik yang digunakan untuk menentukan bahwa penerapan metode yang digunakan telah mampu memecahkan masalah. Pada pengujian ini menggunakan aplikasi SPSS Statistics 24.0.

8. Menarik Kesimpulan

Dalam tahap ini adalah menarik kesimpulan apakah informasi informasi yang didapatkan sesuai dengan fakta atau hipotesis, sehingga akan memperoleh pengetahuan baru.

4. Hasil dan Pembahasan

4.1. Penerapan K-Means Clustering

Atribut data yang digunakan dalam proses ini adalah keluaran tahap FRM yang diambil dari 1 Januari 2017 sampai 30 September 2017 sebanyak 1.030 record.

Tabel 4.1 Tabel RFM telah di transformasi

NO OUTLET	NAMA OUTLET	R	F	M
640000	PENJUALAN KANTOR	18	3	Rp 675.123
640001	PENJUALAN TOPPING U	129	3	Rp 977.719
640002	PENJUALAN NED 1+1	18	3	Rp 640.921

4.1.1. Melakukan Standarisasi Data

Sehingga menghasilkan 3 variabel baru berawalan kata 'z' didepan masing-masing atribut seperti tabel 4.2.

Tabel 4.2 Data Setelah Standarisasi Z-Score

NO OUTLET	NAMA OUTLET	R	F	M	ZR	ZF	ZM
640000	PENJUALAN KANTOR	18	3	Rp675.123	-0,96525	-0,43709	-0,22126
640001	PENJUALAN TOPPING U	129	3	Rp977.719	0,18134	-0,43709	-0,21446
640002	PENJUALAN NED 1+1	18	3	Rp640.921	-0,96525	-0,43709	-0,22203

Sumber: Hasil olah data menggunakan SPSS 24.00 for windows

Tabel 4.3 Output Descriptive Z-Score

Descriptive Statistics					
	N	Min	Max	Mean	Std. Deviation
R	1030	18	271	111,44	96,809
F	1030	1	100	7,14	9,478
M	1030	220058	660446331	10514947,82	44471741,590
Valid N (listwise)	1030				

Sumber: Hasil olah data menggunakan SPSS 24.00 for windows

4.1.2. Analisis Cluster

Untuk menentukan jumlah cluster yang baik maka dilakukanlah uji coba data dimulai dari cluster 2 sampai dengan cluster 5 dilihat dari Jumlah Iterasi dan Analisis Oneway Anova menggunakan aplikasi SPSS dan menghasilkan cluster 3 merupakan pilihan yang paling baik dari jumlah yang lainnya.

Tabel 4.4 Hasil Analisis Pemilihan Jumlah Cluster

Cluster	Jumlah Iterasi	Jumlah Analisis Oneway Anova antar cluster yang tidak signifikan
2	6	Tidak bisa diproses
3	8	2
4	11	6
5	10	14

4.1.2.1. Hasil Output Analisis Cluster

Hasil akhir yang didapatkan setelah proses analysis K-Means cluster adalah sebagai berikut:

Tabel 4.5 Final Cluster Centers

Final Cluster Centers			
	Cluster		
	1	2	3
Zscore(R)	-,69355	1,25927	1,34584
Zscore(F)	-,31837	4,63967	,41108
Zscore(M)	-,16450	6,63710	-,00340

Sumber: Hasil olah data menggunakan SPSS 24.00 for windows

Output final *cluster centers* tersebut masih terkait dengan proses standarisasi data sebelumnya yang mengacu pada *z-score*, dengan ketentuan:

1. Angka negatif (-) berarti data dibawah rata-rata total
2. Angka positif (+) berarti data diatas rata-rata total

Perhitungan tafsiran angka tiap cluster yang terlihat pada tabel 4.6 diatas adalah sebagai berikut :

$$x = \mu + z.\sigma \quad (3)$$

X= rata-rata

μ= standar deviasi

Z= nilai standarisasi yang didapat pada SPSS

Menghitung tafsiran pada variabel RFM tiap-tiap cluster:

X = (rata-rata variabel data pelanggan) + (hasil z-score final variabel x standar deviasi rata-rata variabel data pelanggan).

Menghitung tafsiran pada *Recency*:

Rata-rata R Cluster 1= 44,30338733= 44

Rata-rata R Cluster 2= 233,3529412 = 233

Rata-rata R Cluster 3= 241,7335329 = 242

Menghitung tafsiran pada *Frequency*:

Rata-rata F Cluster 1= 4,125184094 = 4

Rata-rata F Cluster 2= 51,11764706 = 51

Rata-rata F Cluster 3= 11,03892216 = 11

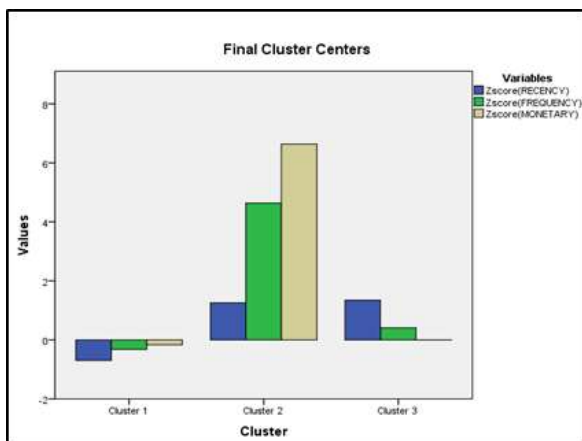
Menghitung tafsiran pada *Monetary*:

Rata-rata M Cluster 1= 3.199.278

Rata-rata M Cluster 2= 305.678.180

Rata-rata M Cluster 3= 10.363.944

Dari hasil cluster diatas dapat ditunjukkan pada gambar 4.1 grafik hasil cluster berikut.



Gambar 4.1 Grafik Final Cluster Centers

Definisi grafik 4.1 ditampilkan pada tabel 4.6.:

Tabel 4.6 Definisi Hasil Akhir

Cluster	R	F	M	Kriteria
1	44 hari ↓ bertransaksi dengan jarak waktu yang dekat	4 kali ↓ pembelian yang paling sedikit	3.199.278 ↓ total pembelian bernilai kecil	pelanggan baru karena baru melakukan pembelian

Tabel 4.6 Definisi Hasil Akhir (lanjutan)

Cluster	R	F	M	Kriteria
2	233 hari ↑ bertransaksi dengan jarak waktu yang cukup lama	51 kali ↑ pembelian yang paling banyak	305.678.180 ↑ total pembelian paling tinggi	pelanggan yang memiliki loyalitas tinggi dan pelanggan yang memberikan nilai kepada perusahaan karena sering melakukan transaksi dan dengan nominal yang besar.
3	242 hari ↑ bertransaksi dengan jarak waktu yang paling lam	11 kali ↓ pembelian yang cukup sedikit	10.363.944 ↓ total pembelian bernilai rendah	pelanggan tidak loyal atau berpotensi tidak aktif karena jarang melakukan transaksi dengan nominal yang kecil.

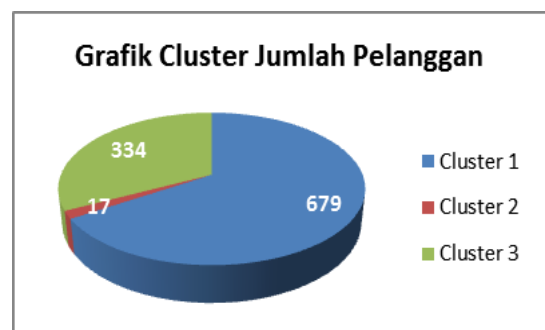
Jumlah Anggota Di Setiap Cluster

Tabel 4.7 Jumlah Anggota Cluster

Number of Cases in each Cluster		
Cluster	1	2
	679,000	
	2	17,000
	3	334,000
Valid		1030,000
Missing		,000

Sumber: Hasil olah data menggunakan SPSS 24.00 for windows

Dari tabel 4.7, dapat dilihat bahwa data cluster pelanggan berjumlah 1030 data dan tidak ada data yang hilang. Pelanggan paling banyak berada di cluster 1 dengan jumlah 679 pelanggan, cluster 2 mempunyai jumlah pelanggan paling sedikit yaitu 17 pelanggan dan cluster 3 mempunyai pelanggan dengan jumlah 334.



Gambar 4.2 Grafik Cluster Jumlah Pelanggan

4.2. Usulan Strategi untuk Perusahaan

Setelah dilakukan pengelompokkan pelanggan menjadi 3 cluster maka dilakukanlah usulan startegi sebagai salah satu referensi bagi perusahaan dalam mengambil kebijakan atau strategi perusahaan dalam mempertahankan atau meningkatkan jumlah pelanggan. Salah satu strategi yang diusulkan adalah strategi pemasaran dan *stategi customer ± market sales promotion techniques*.

1. Strategi promosi penjualan yang dapat dilakukan perusahaan yaitu:

- a. Customer promotion

Promosi yang bermaksud membujuk konsumen dalam melakukan pembelian terhadap suatu produk barang. Seperti pemberian hadiah gantungan kunci, pulpen, gelas dan harga spesial untuk pelanggan yang melakukan pre-order.

- b. Sales-force promotion

Langkah promo yang bertujuan untuk memberi motivasi pada armada penjualan yang dimiliki perusahaan.

- c. Bussiness promotion

Promo penjualan untuk mendapatkan pelanggan baru, menjaga interaksi dan komunikasi dengan pelanggan, mensosialisasikan produk baru, serta melakukan penjualan lebih banyak pada pelanggan tetap.

2. Strategi customer $\pm$ market sales promotion techniques adalah sebagai berikut:

- a. Kupon

Sertifikat dengan nilai tertera untuk pengurangan harga pada barang-barang tertentu atau memberikan hak kepada pemegangnya untuk mendapatkan potongan harga seperti tercetak untuk pembelian produk dengan jumlah tertentu.

- b. Sampling

Penawaran kepada calon pelanggan untuk berkesempatan mencoba produk sebelum mengambil keputusan membeli

- c. Price-Off Deals

Memberikan potongan harga langsung ditempat pembelian.

- d. Content and Sweepstakes

Hadiah seperti tawaran kesempatan untuk memenangkan emas murni, uang tunai, voucher liburan, voucher belanja atau barang-barang karena mempunyai nilai monetary yang tinggi.

- e. *Frequency*

Teknik yang mengarah kepada program-program yang berkelanjutan seperti menawarkan konsumen diskon atau hadiah langsung gratis untuk mencapai terjadinya pengulangan dalam pembelian atau langganan.

5. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan maka didapatkan beberapa kesimpulan sebagai berikut:

1. Setelah dilakukan *K-Means Clustering* data pelanggan menggunakan RFM yang dibagi menjadi tiga *cluster* maka menghasilkan karakteristik pelanggan yang dijelaskan pada tabel 5.1.

Tabel 5.1 Hasil Pengelompokkan Pelanggan

Cluster	Jumlah Pelanggan	Kriteria Recency	Kriteria Frequency	Kriteria Monetary	Kategori Pelanggan	Usulan Strategi Promosi	Usulan Strategi Customer Interact Sales Promotion Techniques
1	879	+ Rendah (lama belum beli)	+ Rendah (sedikit transaksi)	+ Rendah (Total biaya pembelian sedikit)	Pelanggan baru (new customer)	Customer promotion, Sales force promotion, Business promotion	Sampling, Frequency, Price-Off Deal
2	17	+ Tinggi (lama belum beli)	+ Tinggi (banyak transaksi)	+ Tinggi (Total biaya pembelian banyak)	Pelanggan Loyal Utama	Customer promotion, Sales force promotion	Kupon, Frequency, Content and Sweepstakes
3	334	+ Tinggi (lama belum beli)	+ Rendah (sedikit transaksi)	+ Rendah (Total biaya pembelian sedikit)	Pelanggan tidak loyal (disengaja tidak beli)	Customer promotion, Sales force promotion	Price-Off Deal, Frequency

2. Berdasarkan analisis ANOVA dihasilkan variabel yang sangat membedakan karakteristik dari keempat cluster adalah faktor *Recency*. Berdasarkan analisis Oneway Anova maka dihasilkan variabel frequency dan monetary mempunyai perbedaan yang signifikan terhadap semua cluster. Untuk variabel *recency*, perbedaan antara cluster 2 terhadap 3, cluster 3 terhadap 2 mempunyai perbedaan yang tidak signifikan karena mempunyai tetapi masih memiliki perbedaan walaupun kecil.

Daftar Pustaka

- Amborowati, A., & Winarko, E., 2014. Review Pemanfaatan Teknik Data Mining Dalam Segmentasi Konsumen. KOMMIT Vol 8, 66-73.
- Astria, D., & Suprayogi., 2017. Penerapan Algoritma Fuzzy C-Means untuk Clustering Pelanggan Pada CV Mataram Jaya Bawean. Eksplora Informatika Vol 6 No 2 , 169-178.
- Chandra, K., 2013. Clustering Nasabah Bank Menggunakan K-Means Berdasarkan Nilai RFM (Recency,Frequency, Monetary). Media Informatika Vol 12 No 3, 39-55.
- Hardiani, T., Sulisty, S., & Hartanto, R., 2015. Segmentasi Nasabah Tabry) dan K-Means pada Lembaga Keuangan Mikroungan Menggunakan Model RFM (Recency, Frequency, Monet. SEMANTIK, 463-468.
- Herwin, Handhayani, T., & Herwindiati, D. E., 2015. Sistem Prediksi Customer Loyalty dengan Metode Segmentasi RFM dan Pengelompokan dengan Algoritma Clustering Modified K-Means pada Perusahaan PU IBUKOTA. Seminar Nasional Teknologi Informasi, 32-38.
- Kristanto, T., & Arief, R., 2013. Analisis Data Mining Metode Fuzzy untuk Customer Relationship Management pada Perusahaan Tour & Travel. Seminar Nasional Sistem Informasi Indonesia, 528-534.
- Kusrini, & Luthfi, E. T. , 2009. Algoritma Data Mining. Yogyakarta: Andi Offset.
- Ong, J. O., 2013. Implementasi Algoritma K-Means Clustering untuk menentukan Strategi Marketing President University. Jurnal Ilmiah Teknik Industri Vol 12 No 1, 10-20.
- Pramesti, G., 2014. Kupas Tuntas Data Penelitian dengan SPSS 22. Jakarta: PT Alex Media Komputindo.
- Rajagopal, S., 2011. CUSTOMER DATA CLUSTERING USING DATA. International Journal of Database Management Systems (IJDMS) Vol 3 No 4, 3-11.
- Rumiarti, C. D., & Budi, I., 2017. Segmentasi Pelanggan pada Customer Relationship Management di Perusahaan Ritel: Studi Kasus PT Gramedia Asri Media. Jurnal Sistem Informasi Vol 1 No 13, 1-10.
- Setiawan, I., 2015. Pengelompokan Nasabah Bank menggunakan Algoritma K-Means untuk memberikan Penawaran yang Tepat. Seminar Nasional Teknik Informatika dan Multimedia , 79-84.
- Silwattananusarn, T., & Tuamsuk, K., 2012. Data Mining and Its Applications for Knowledge. International Journal of Data Mining & Knowledge Management Process (IJDMP) Vol2 No5, 13-24.
- Sudriyanto., 2017. Clustering Loyalitas Pelanggan dengan Metode RFM (Recency, Frequency, Monetary) dan Fuzzy C-Means. Prosiding SNATIF ke 4, 815-822.
- Sumadikarta, I., & Abeiza, E. (t.thn.). Penerapan Algoritma K-Means Pada Data Mining Untuk Memilih Produk dan Pelanggan Potensial. Jurnal Styta Informatika Vol 1 No 1, 12-22.
- Susanto, S., 2015. Menguasai Statistik Multivariat. Jakarta: PT Alex Media Komputindo.