

Analisis Segmentasi Agen Penjualan Menggunakan Model RFM dan Algoritma *K-means*

Adee Purnama Jaya^{a1}, Adi Hartanto, ST, M.Kom^{a2}

^aUniversitas Mercu Buana, Fakultas Ilmu komputer
Jakarta, Indonesia

¹adee.purnama.jaya@gmail.com

²adi.hartanto@mercubuana.ac.id

Abstrak

Agen penjualan memberikan kontribusi yang cukup besar dalam pemasaran. Setiap agen yang berbeda mempunyai nilai yang berbeda pula. Pengetahuan tentang karakteristik setiap agen diperlukan untuk mendukung keputusan terkait strategi bisnis perusahaan serta mengelola hubungan yang baik Antara perusahaan dan agennya. Segmentasi agen penjualan dapat dilakukan untuk mendapatkan pengetahuan tentang karakteristik setiap agen. Dengan segmentasi agen penjualan, perusahaan diharapkan dapat menerapkan kebijakan yang tepat. Informasi tentang segmentasi agen penjualan dapat diperoleh dengan pendekatan teknik *Data Mining* dan metode RFM (*Recency, Frequency, Monetary*). Teknik *Data Mining K-means Clustering* digunakan untuk segmentasi agen penjualan. RFM adalah model yang digunakan untuk membedakan agen berdasarkan 3 variabel, yaitu kebaruan, frekuensi, dan moneter. Dari hasil penerapan model k-means dan RFM agen penjualan dibagi menjadi 3 segmen dengan karakteristik yang berbeda. Ketiga segmen tersebut yaitu 54.17% agen baru atau agen yang jarang terjadi transaksi dan belum lama terjadi transaksi lagi, 35.42% agen penjualan terbaik dan merupakan agen yang "bernilai" bagi perusahaan, serta 10.42% agen penjualan yang aktifitasnya menurun.

Kata Kunci: *Data Mining, k-means, Clustering, RFM.*

1. Pendahuluan

Persaingan dalam dunia bisnis yang dinamis dan kompetitif mengharuskan perusahaan untuk melakukan strategi bisnis yang inovatif dan tepat sasaran. Pertumbuhan bisnis yang disertai dengan kemajuan dan perkembangan teknologi, juga memerlukan strategi bisnis yang tepat sehingga bisnis dapat terus berjalan dan berkembang. Berkembangnya suatu bisnis berbanding lurus dengan tingkat profit perusahaan. Strategi bisnis yang dapat dilakukan untuk meningkatkan profitabilitas, pendapatan dan kepuasan pelanggan pada tahap awal adalah mengatur segmentasi [1]. Agen merupakan salah satu asset penting dalam pemasaran. Segmentasi agen penjualan penting karena setiap agen memiliki karakteristik dan perilaku yang berbeda [1]. Oleh karena itu, diperlukan strategi bisnis yang berbeda. Strategi bisnis yang tepat dapat diterapkan jika ada pengetahuan tentang karakteristik agen penjualan. Informasi tentang segmentasi agen penjualan dapat diperoleh dengan pendekatan teknik *Data Mining* dan metode RFM.

Data Mining merujuk pada penggalian informasi dari sejumlah besar data, dan mengubah informasi itu menjadi struktur yang dapat dipahami dan bermakna untuk digunakan lebih lanjut [2] Teknik utama untuk menemukan pola adalah asosiasi-aturan / penambangan pola-sering, *clustering*, klasifikasi dan analisis regresi [3]. *Clustering* adalah proses pengelompokan satu set objek sedemikian rupa sehingga objek dalam grup yang sama lebih mirip dalam beberapa cara satu sama lain daripada pada kelompok lain [4]. *Clustering* adalah tugas penting dalam analisis data dan aplikasi *Data Mining* [2]. *K-means* adalah salah satu algoritma yang terkenal untuk *clustering* dan telah digunakan secara luas di berbagai bidang termasuk *Data Mining*, data

statistik, analisis dan aplikasi bisnis lainnya [5]. K-means adalah algoritma pembelajaran unsupervised yang paling mudah dan populer yang memecahkan masalah *clustering* populer [6]. jika menggunakan metode k-means, harus dilakukan transmisi data agar data dapat diolah. Data yang berjenis data nominal seperti tanggal transaksi dan kode pemesanan harus diinisialisasikan terlebih dahulu kedalam bentuk angka. Jika data sudah dalam bentuk angka, barulah data dapat diproses

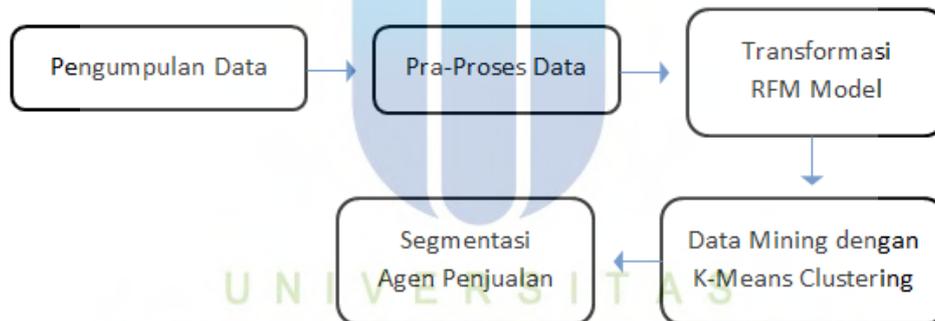
Model RFM merupakan model yang membedakan pelanggan penting dari data besar oleh tiga variabel yaitu recency, frequency dan monetary[7]. Variabel atau Indikator dalam analisis RFM adalah [1]:

- Recency dari pembelian terakhir (R); R merepresentasikan kebaruan, mengacu pada interval waktu pembelian terakhir hingga waktu saat ini.
- Frequency pembelian (F); F menunjukkan frekuensi mengacu pada jumlah transaksi dalam periode waktu tertentu.
- Nilai Monetary dari pembelian (M); M merepresentasikan Monetary mengacu pada berapa banyak uang yang dikonsumsi (dihabiskan/digunakan) dalam periode waktu tertentu.

Pada penelitian ini, teknik *Data Mining k-means clustering* dipadukan dengan model RFM dieksplorasi untuk segmentasi agen penjualan. Dengan implementasi *k-meansi*, pengetahuan tentang karakteristik setiap agen dapat diketahui.

2. Metode Penelitian

Metode penelitian yang digunakan dalam penerapan algoritma *K-means* untuk menentukan segmentasi agen penjualan tiket kereta api, menggunakan rancangan penelitian yang tahapan proses penelitian ditunjukkan pada Gambar 2.1 berikut:



Gambar 2.1: Tahapan Proses Penelitian

Dari gambar 2.1 terlihat ada 5 tahapan proses penelitian yaitu tahap Pengumpulan data, Pra-proses data, transformasi data ke model RFM, *Data Mining* dengan *K-means Clustering*, dan terakhir segmentasi agen penjualan.

2.1 Dataset

2.1.1 Sumber

Data yang digunakan dalam penelitian ini diperoleh melalui aplikasi "*railticket*" PT.KAI yang diakses dan diunduh pada tanggal 03 Desember 2018.

2.1.2 Data Analisis

Data yang digunakan dalam penelitian ini adalah data transaksi penjualan dan pemesanan tiket Kereta Api Pangrango dengan nomor KA 391 pada bulan November 2017. Data tersebut adalah data penjualan dan pemesanan tiket secara online maupun offline melalui loket penjualan langsung di stasiun kereta.

Tabel 2.1: Cuplikan data penjualan dan pemesanan tiket

STASIUN JUAL	TGL BAYAR	KODE PEMESANAN	JUMLAH
ALFAMIDI	01-NOV-17	8ADG1F	15000
ATA	01-NOV-17	E7YSDR	15000
ATA	01-NOV-17	SS5CGJ	20000
ATA	01-NOV-17	JXWMUE	20000
ATA	01-NOV-17	NEVBPE	20000
BLIBLI.COM	01-NOV-17	ZT1UPG	15000
BMW	01-NOV-17	SEAP3D	20000
BMW	01-NOV-17	8EF6Y1	15000
BMW	01-NOV-17	6RW46Y	20000
BMWWEBIN	01-NOV-17	XESTEA	60000

2.2 Pra-Proses Data

Pada tahap ini data yang akan digunakan dipersiapkan untuk mempermudah proses *Data Mining*. Ada beberapa proses yang dilakukan dapat mempersiapkan data diantaranya memilih variabel yang akan dianalisis, membersihkan data, menyiapkan data awal sehingga siap untuk data transformation [7].

Data pada tabel 2.1 untuk dapat diolah dengan menggunakan algoritma *k-means* data tersebut perlu diinisialisasikan kedalam bentuk angka. Segmentasi agen penjualan pada penelitian ini menggunakan model RFM (*Recency, Frequency, Monetary*). Untuk mengubah dataset kedalam model RFM maka perlu dilakukan seleksi data dari model RFM tersebut yaitu:

- R (Recency)* adalah rentang antar waktu terakhir melakukan transaksi sampai tanggal yang ditentukan, pada penelitian ini penulis menentukan pada tanggal 30 November 2017,
- F (Frequency)* adalah jumlah transaksi yang terjadi,
- M (Monetary)* adalah jumlah nominal transaksi dalam rupiah.

Tabel 2.2 berikut menunjukkan hasil dari seleksi data.

Tabel 2.2: Seleksi data

Data Awal	Data Akhir
Tanggal terakhir transaksi (tipe: <i>date</i>)	<i>Recency</i> (tipe: <i>number</i>)
Jumlah transaksi (tipe: number)	<i>Frequency</i> (tipe: <i>number</i>)
Jumlah nominal transaksi (tipe: number)	<i>Monetary</i> (tipe: <i>number</i>)

2.3 Transformasi RFM Model

Dalam tahap proses transformasi RFM Model, seperti yang dijelaskan pada gambar 2.1. Hasil yang di dapat adalah seperti pada tabel 2.3 berikut:

Tabel 2.3: Nilai RFM agen penjualan

NO	STASIUN JUAL	R	F	M
1	ALFAMIDI	0	77	2135000
2	ATA	0	103	2630000
3	BD	3	1	50000
4	BKS	2	4	225000
5	BLIBLI.COM	1	26	635000
6	BMW	0	59	1720000
7	BMWEBIN	1	46	1275000
8	BOO	0	162	6005000
9	BTT	2	16	500000
10	BUKALAPAK	0	67	1810000

Pada tabel 2.3 atribut STASIUN JUAL berisi nama nama agen penjualan, *RECENCY* berisi jumlah hari terakhir transaksi dari tanggal 30 november 2017, *FREQUENCY* berisi jumlah transaksi, dan *MONETARY* berisi jumlah nilai transaksi dalam rupiah.

Setelah nilai RFM untuk setiap Agen penjualan didapat. Nilai RFM dikelompokkan menjadi 5 bagian. Tabel 2.4 menunjukkan aturan pengelompokan.

Tabel 2.4: Aturan penilaian (Scaling) RFM

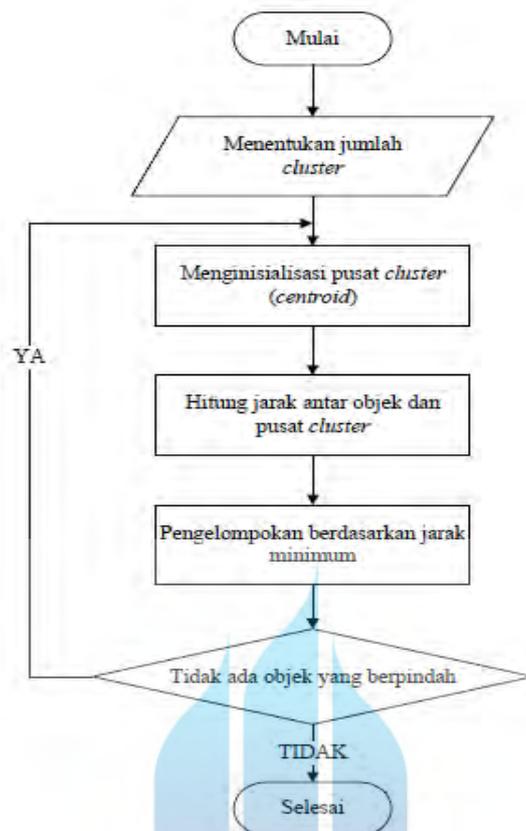
SCORE	R	F	M
5	<5	>500	>10.000.000
4	<10	<500	<10.000.000
3	<15	<250	<5.000.000
2	<20	<100	<2.500.000
1	>20	<50	<1000.000

Tabel 2.5: Nilai RFM setelah penilaian (Scaling)

NO	STASIUN JUAL	R	F	M
1	ALFAMIDI	5	2	2
2	ATA	5	3	3
3	BD	5	1	1
4	BKS	5	1	1
5	BLIBLI.COM	5	1	1
6	BMW	5	2	2
7	BMWEBIN	5	1	2
8	BOO	5	3	4
9	BTT	5	1	1
10	BUKALAPAK	5	2	2

2.4 Penerapan Algoritma *K-means*

Tahap proses segmentasi agen penjualan tiket, pada penelitian ini menggunakan metode *K-means Clustering*. Metode *K-means Clustering* dapat dilihat pada Gambar 2.2 berikut :



Gambar 2.2: Flowchart Algoritma K-means

Dari gambar diatas, dapat diketahui langkah-langkah dalam melakukan *clustering* dengan metode *K-means*, adalah sebagai berikut :

- Tentukan jumlah cluster yang akan dibentuk
- Tentukan titik pusat awal (Centroid) dari setiap cluster
- Tentukan jarak objek ke Centroid menggunakan rumus euclidean distance, dengan rumus sebagai berikut : rumus[1]

$$D(ij) = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + \dots + (Xki - Xkj)^2} \quad (1)$$

- Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid (jarak terkecil).
- Hitung kembali nilai centroid. Nilai centroid baru diperoleh dari rata-rata cluster dengan keanggotaan yang sekarang.
- Lakukan kembali langkah ke-2 menggunakan nilai centroid yang baru. Jika nilai centroid tidak berubah lagi maka proses selesai

3. Hasil Dan Pembahasan

Pada proses sebelumnya, data yang diperoleh telah diproses dengan metode RFM dan nilai-nilainya dikelompokkan menjadi 5 kelompok seperti pada tabel 3 diatas. Data sampel yang diproses berjumlah 48 baris dengan 3 atribut RFM.

3.1. Pengolahan Data Manual

Data sampel kemudian dilakukan pengolahan data dengan proses *clustering* dengan menggunakan algoritma *k-means* sehingga diperoleh hasil pengelompokan data yang diinginkan. Adapun langkah-langkah yang dilakukan adalah sebagai berikut:

3.1.1. Menentukan jumlah *cluster*

Penulis menettukan jumlah *cluster* yang akan dibentuk sebanyak 3 *cluster*. Diharapkan dari ketiga *cluster* tersebut diperoleh kelompok agen penjualan dengan karakteristik terbaik, kurang baik, ataupun agen yang mengalami penurunan aktifitas.

3.1.2. Menentukan titik pusat awal (*centroid*)

Penentuan titik *centroid* ditentukan secara acak diambil dari data yang ada. Nilai *cluster* k_0 diambil dari baris ke-25, nilai *cluster* k_1 diambil dari baris ke-22, dan nilai *cluster* k_2 diambil dari baris ke-9.

Tabel 3.1: Nilai titik pusat awal (*centroid*)

<i>Cluster</i>	R	F	M
k_0	5	5	5
k_1	1	1	1
k_2	5	1	1

3.1.3. Menghitung jarak dari *centroid*

Untuk menghitung jarak antara titik *centroid* dengan titik tiap objek menggunakan rumus *euclian distance*. Berikut adalah penghitungan jarak *euclian distance* secara manual :

a. Jarak antara data ke-1 dengan *centroid cluster* k_0

$$D(1.0) = \sqrt{(5 - 5)^2 + (2 - 5)^2 + (2 - 5)^2} = 4.2426$$

b. Jarak antara data ke-1 dengan *centroid cluster* k_1

$$D(1.1) = \sqrt{(5 - 1)^2 + (2 - 1)^2 + (2 - 1)^2} = 4.2426$$

c. Jarak antara data ke-1 dengan *centroid cluster* k_2

$$D(1.2) = \sqrt{(5 - 5)^2 + (2 - 1)^2 + (2 - 1)^2} = 1.4142$$

Tabel 3.2: Hasil perhitungan iterasi ke-1

STASIUN JUAL	D_0	D_1	D_2	k_0	k_1	k_2
ALFAMIDI	4.2426	4.2426	1.4142	0	0	1
ATA	2.8284	4.8990	2.8284	1	0	1
BD	5.6569	4.0000	0.0000	0	0	1
BKS	5.6569	4.0000	0.0000	0	0	1
BLIBLI.COM	5.6569	4.0000	0.0000	0	0	1
BMW	4.2426	4.2426	1.4142	0	0	1
BMWEBIN	5.0000	4.1231	1.0000	0	0	1
BOO	2.2361	5.3852	3.6056	1	0	0
BTT	5.6569	4.0000	0.0000	0	0	1
BUKALAPAK	4.2426	4.2426	1.4142	0	0	1

Dari hasil perhitungan diatas didapatkan titik pusat *centroid* baru seperti pada table berikut:

Tabel 3.3: Nilai *centroid* iterasi ke-1

<i>Cluster</i>	R	F	M
k_0	5.0000	3.8235	4.2941
k_1	2.2000	1.0000	1.0000
k_2	4.7185	1.4063	1.5313

Iterasi ke-2 dilakukan perhitungan jarak *euclian distance* dengan nilai *centroid* baru hasil iterasi ke-1. Dari hasil perhitungan iterasi ke-2 didapatkan titik pusat *centroid* baru seperti pada table 3.4 berikut:

Tabel 3.4: Nilai *centroid* iterasi ke-2

Cluster	R	F	M
<i>k0</i>	5.0000	3.8235	4.2941
<i>k1</i>	2.2000	1.0000	1.0000
<i>k2</i>	4.8077	1.1923	1.3462

Iterasi selanjutnya melakukan perhitungan jarak *euclian distance* dengan nilai *centroid* baru hasil iterasi ke-2

Tabel 3.5: Hasil perhitungan iterasi ke-3

STASIUN JUAL	<i>D0</i>	<i>D1</i>	<i>D2</i>	<i>k0</i>	<i>k1</i>	<i>k2</i>
ALFAMIDI	2.9306	3.1369	1.0568	0	0	1
ATA	1.5339	3.9799	2.4576	1	0	0
BD	4.3386	2.8000	0.4402	0	0	1
BKS	4.3386	2.8000	0.4402	0	0	1
BLIBLI.COM	4.3386	2.8000	0.4402	0	0	1
BMW	2.9306	3.1369	1.0568	0	0	1
BMWEBIN	3.6380	2.9732	0.7082	0	0	1
BOO	0.8745	4.5651	3.2168	1	0	0
BTT	4.3386	2.8000	0.4402	0	0	1
BUKALAPAK	2.9306	3.1369	1.0568	0	0	1
CBD	0.8745	4.5651	3.2168	1	0	0
CC	3.6380	2.9732	0.7082	0	0	1
CCKP	4.3386	2.8000	0.4402	0	0	1
CCR	1.5339	3.9799	2.4576	1	0	0
CGB	4.3386	2.8000	0.4402	0	0	1

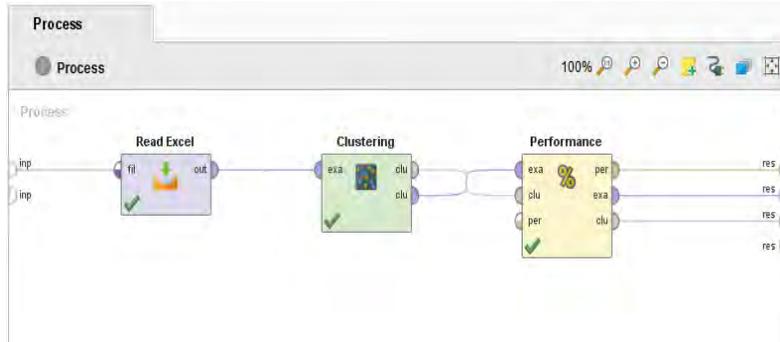
Karena pada itersi ke-2 dan ke-3 posisi *cluster* tidak berubah maka iterasi dihentikan dan hasil akhir yang diperoleh yaitu:

- Cluster k0* memiliki 17 data
- Cluster k1* memiliki 5 data
- Cluster k2* memiliki 26 data

3.2. Pengolahan Data Menggunakan Rapidminer

Pada tahap implementasi dan pengujian, penelitian ini menggunakan *software* rapidminer Studio Version 9.0.003 yang berguna dalam membantu melakukan kegiatan *Data Mining*, Klustering, klasifikasi, dan regresi. RapidMiner merupakan *software/perangkat lunak* untuk pengolahan data. Dengan menggunakan prinsip dan algoritma *Data Mining*, RapidMiner mengekstrak pola-pola dari data set yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan database [8].

Dengan impelentasi dan pengujian menggunakan *software* penulis akan membandingkan bagaimana hasil pengolahan secara manual dengan hasil pengolahan data menggunakan *software* Rapidminer. Berikut adalah tampilan *Process clustering* menggunakan algoritma *k-means* pada Rapidminer :



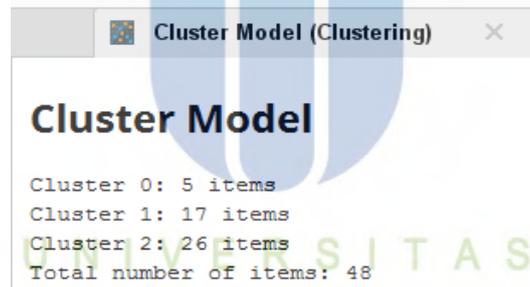
Gambar 3.1: halaman process *clustering* rapidminer

Dari Gambar 3.1 dijelaskan sebagai berikut:

- Read Excel : yaitu operator ini membaca ExampleSet dari file Excel yang ditentukan
- Clustering* : Adalah Operator yang melakukan pengelompokan menggunakan algoritma *K-means*
- Performance : adalah operator yang digunakan untuk evaluasi kinerja metode pengelompokan berbasis centroid. Operator ini memberikan daftar nilai kriteria kinerja berdasarkan centroid cluster.

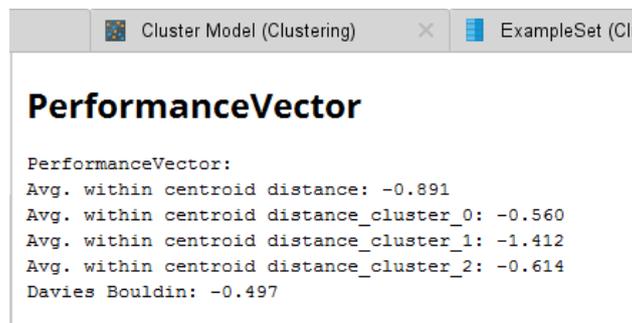
Proses *clustering* dan evaluasi cluster menggunakan software rapidminer studio. Davies bouldin index digunakan untuk mengidentifikasi nilai cluster k yang optimal. Semakin kecil nilai davies bouldin index semakin optimal hasil *clustering*.

Hasil proses *clustering* menggunakan software rapidminer dengan 3 cluster yang ditunjukkan pada gambar 3.2 berikut:



Gambar 3.2: Hasil cluster model

Selain itu hasil proses *clustering* menggunakan software Rapidminer didapat nilai *Davies Bouldin Index* (DBI). *Davies bouldin index* digunakan untuk mengidentifikasi nilai cluster k yang optimal/evaluasi *cluster*. Semakin kecil nilai davies bouldin index semakin optimal hasil *clustering*. Nilai hasil evaluasi *cluster* pada tahap implementasi dan pengujian ini adalah -0.479 yang nilainya baik, yang ditunjukkan pada gambar 3.3 berikut:



Gambar 3.3: Nilai DBI

Pada tabel dibawah ini ditampilkan perbandingan hasil pengolahan data secara manual dan menggunakan *software* rapidminer pada proses *clustering* menggunakan algoritma *k-means*.

Tabel 3.6: Perbandingan keanggotaan segmen

<i>Attribute</i>	<i>k0/cluster_0</i>	<i>k1/cluster_1</i>	<i>k2/cluster)2</i>
Manual	17	5	26
Rapidminer	5	17	26

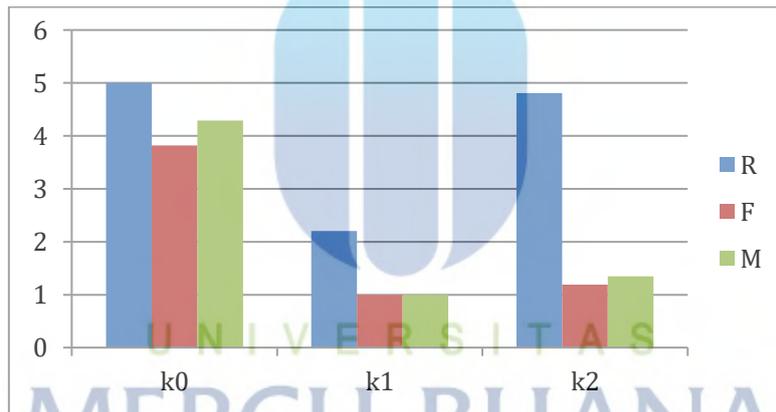
Tabel 3.7: Perbandingan nilai centroid akhir

Nilai *centroid* hasil *clustering* manual

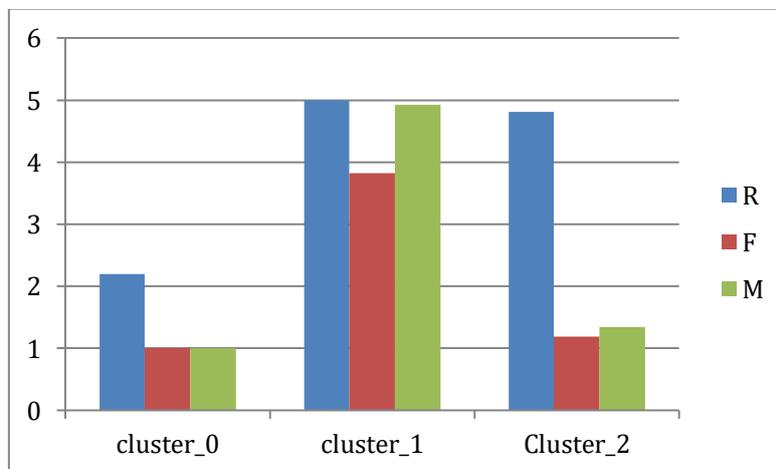
<i>Attribute</i>	<i>k0</i>	<i>k1</i>	<i>k2</i>
R	5.0000	2.2000	4.8077
F	3.8235	1.0000	1.1923
M	4.2941	1.0000	1.3462

Nilai *centroid* hasil *clustering* rapidminer

<i>Attribute</i>	<i>cluster_0</i>	<i>cluster_1</i>	<i>cluster_2</i>
R	2.2000	5.0000	4.8080
F	1.0000	3.8240	1.1920
M	1.0000	4.9240	1.3460



Gambar 3.4: Diagram nilai *centroid* hasil manual



Gambar 3.5: Diagram nilai *centroid* hasil rapidminer

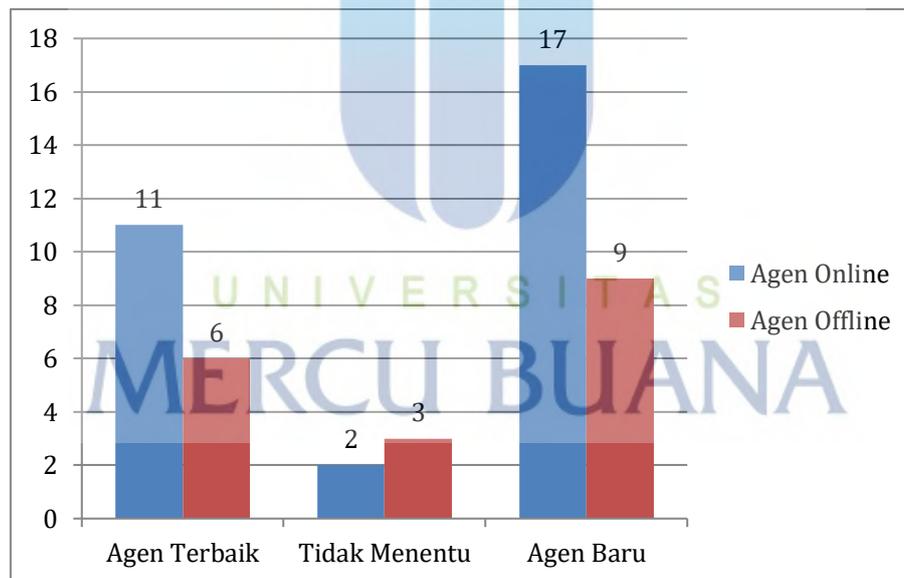
Dari kedua diagram nilai *centroid* akhir terlihat keduanya menghasilkan segmentasi data dengan karakteristik yang sama. Segmentasi data yang dihasilkan yaitu:

- Segmen 1 adalah cluster dengan nilai *Recency* (R), *Frequency* (F), dan *Monetary* (M) yang tinggi. Agen yang berada pada cluster ini adalah agen terbaik perusahaan (Terbaik).
- Segmen 2 adalah cluster dengan nilai *Recency* (R), *Frequency* (F), dan *Monetary* (M) yang rendah. Agen yang berada pada cluster ini adalah agen yang memiliki kontribusi yang rendah terhadap perusahaan (Tidak Menentu).
- Segmen 3 adalah cluster dengan nilai *Recency* tinggi, *Frequency* dan *Monetary* yang rendah. Agen yang berada pada cluster ini diklasifikasikan sebagai agen baru atau agen yang jarang terjadi transaksi dan belum lama terjadi transaksi lagi (Agen Baru).

Ketiga segmen diatas masing-masing memiliki keanggotaan agen online dan agen offline seperti yang ditunjukkan pada tabel dibawah ini:

Tabel 3.8: Keanggotaan Segmen

	Segemen 1 (Agen Terbaik)	Segmen 2 (Tidak Menentu)	Segmen 3 (Agen Baru)	Jumlah	%
Agen Online	11	2	17	30	62.5
Agen Offline	6	3	9	18	37.5
Jumlah	17	5	26	48	
%	35.42	10.42	54.17		



Gambar 3.6: Keanggotaan Agen Online dan Offline

Dari gambar diatas diketahui perbandingan antara agen online dengan offline dari setiap segmen, yaitu dengan peresentase: segmen 1 = 64.70% : 35.29%, segmen 2 = 40% : 60%, dan segmen 3 = 65.38% : 34.61%.

4. Kesimpulan

Pada penelitian ini dari hasil implementasi dan pengujian yang telah dilakukan sebelumnya diperoleh kesimpulan, metode *Clustering* dengan algoritma *k-means* dapat digunakan untuk mengelompokan data agen penjualan berdasarkan karakteristiknya dengan metode RFM. Hasil dari segmentasi agen penjualan didominasi oleh karakteristik agen baru. 54.17% agen baru

atau agen yang jarang terjadi transaksi dan belum lama terjadi transaksi lagi, sehingga memiliki nilai *Frequency* dan *Monetary* yang rendah. Perusahaan memiliki 35.42% agen penjualan terbaik dan merupakan agen yang "bernilai" bagi perusahaan, serta 10.42% agen penjualan yang aktifitasnya menurun. Perbandingan antara agen online dan offline di setiap segemen dengan berbanding agen secara keseluruhan hampir sama yaitu berada diantara 60-70% berbanding 30-40%. Hal ini berarti dominasi dan komposisi penyebaran agen online dan offline terbagi di setiap segmen rata-rata sama. Dari hasil penelitian dan informasi yang diperoleh, perusahaan diharapkan untuk dapat menetapkan kebijakan yang lebih tepat.

Karena keterbatasan waktu dan pengetahuan, penulis menyadari adanya kekurangan dalam penulisan ini. Untuk memperbaiki dan penyempurnaan penelitian ini penulis memberikan beberapa saran yang dapat dilakukan yaitu:

- a. Untuk memperoleh informasi terbaik dan terupdate dapat menggunakan data yang lebih banyak, parameter yang berbeda, serta data terbaru.
- b. Penentuan nilai *centroid* awal yang baik akan membuat proses *clustering* dapat dilakukan lebih cepat
- c. Pengekstrasian data penjualan dengan *Data Mining* dapat menggunakan metode dan algoritma yang berbeda.

Referensi

- [1] Y. Primawati, I. Verdian, and G. W. Nurcahyo, "Classification of Sales Agent for Cement Distribution using K-Means *Clustering*," *J. of Comput. Sci. and IT.*, Vol. 1, NO. 1, pp. 1-9, 2016.
- [2] K. Chitra, and Dr. D. Maheswari, "A Comparative Study of Various *Clustering* Algorithm in *Data Mining*," *Int. J. of Comput. Sci and Mobile Computing*, vol. 6, issue 8, pp. 109-115, 2017.
- [3] A.S. Hashmi and T. Ahmad, "Big *Data Mining*: Tools Algorithm" *iJES*, vol 4, issue 1, pp. 2016.
- [4] S.S. Ghuman, "*Clustering* Techniques- A Review", *Int. J. Of Comput. Sci and Mobile Computing*, vol. 5, issue .5, pp., 2016.
- [5] H. Tikaridha, S. Selo, and H. Rudy, "Segmentasi Nasabah Tabungan Menggunakan Model RFM (*Recency, Frequency, Monetary*) dan K-Means Pada Lembaga Keuangan Mikro," *SEMANTIK*, 2015.
- [6] Saklecha A, and Raikhwaj J, "Enhanced K-Means *Clustering* Algorithm Using Collaborative Filtering Approach," *Oriental J. Of Comput. Sci, & Tech.*, vol.10, n0.(2), pp. 474-479, 2017.
- [7] B E Andiana, I. Soesanti, and A E Permanasari, " Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM dan Teknik *Clustering*" *JUTEI*, vol.2, no. 1, pp. 2018.
- [8] B. Rahmat C.T.I., A.A. Gafar, N Fajriani, dkk, "Implementasi K-Means *Clustering* Pada Rapidminer Intuk Analisis Daerah Rawan Kecelakaan", *Seminar Nasional Riset Kuantitatif Terapan*, 2017.