K-Means *Clustering* Untuk Segmentasi Produk Berdasarkan Analisis *Recency, Recency, Recency* (RFM) pada Data Transaksi Penjualan

Maulidah Fithriyah^{1,*}, Muhammad Ainul Yaqin², Syahiduz Zaman³

Program Studi Teknik Informatika, Universitas Islam Negeri Maulana Malik Ibrahim, Indonesia

¹14650030@student.uin-malang.ac.id; ²yaqinov@ti.uin-malang.ac.id; ³syahid@ti.uin-malang.ac.id

* corresponding author

INFO ARTIKEL

Sejarah Artikel

Diterima: 29 Juni 2021

Direvisi: 5 Agustus 2021

Diterbitkan: 31 Agustus 2021

Kata Kunci

Product Segmentation

K-Means

Clustering

RFM

ABSTRAK

Pada bisnis *online* maupun *clustering* seringkali ditemukan kendala pada proses pengelolaan stok produk. Oleh karena itu, diperlukan adanya strategi yang tepat untuk mengetahui kebutuhan dan permintaan pelanggan pada suatu produk. Salah satunya adalah dengan melakukan segmentasi produk yang memiliki karakter berbeda sehingga dapat memudahkan pemilik bisnis untuk mengetahui produk yang banyak diminati *clustering*, meminimalisir kekurangan maupun kelebihan stok, dan pemenuhan permintaan *clustering* dapat disiapkan tepat waktu. Pada penelitian ini variabel dari model RFM digunakan dalam analisis data transaksi produk untuk mendapatkan nilai indeks dari masing-masing produk. Semua data produk yang ada dalam data transaksi produk di kelompokkan menggunakan algoritma k-means *clustering* berdasarkan nilai indeksnya. Proses *clustering* dilakukan sebanyak 5 kali dengan jumlah *cluster* 2, 3, 4, 5, dan 6 dengan tujuan untuk mendapatkan jumlah *cluster* yang paling optimal dalam mengelompokkan data produk yang ada. Selanjutnya, hasil masing-masing proses *clustering* dievaluasi dengan menggunakan rumus *silhouette* untuk mengetahui kualitas dari hasil *clustering*. Dari hasil pengujian didapatkan nilai *silhouette* tertinggi adalah 0,4314 pada *clustering* ke 2 dengan jumlah 3 *cluster*. Dari penelitian ini didapatkan 3 kelompok produk dengan karakter yang berbeda. C1 berupa kelompok produk dengan nilai *recency*, *recency*, dan *recency* paling baik. Pada C2 berisi kelompok produk yang memiliki nilai *recency*, *recency*, dan nilai *recency* paling buruk. Pada C3 berisi kelompok produk yang memiliki nilai *recency*, *recency*, dan *recency* yang cukup baik

PENDAHULUAN

Kegiatan jual beli *clustering* maupun *online* seringkali ditemukan kendala pada proses pengelolaan stok produk. Beberapa diantaranya, tidak seimbang antara persediaan produk dengan permintaan pelanggan [1]. Ada masa dimana persediaan produk terbatas sedangkan permintaan pelanggan tinggi, hal tersebut bisa mempengaruhi kepercayaan pelanggan karena pelayanan yang tidak maksimal dalam memenuhi permintaan mereka. Selain itu, jika stok yang tersedia melebihi permintaan pelanggan maka akan menyebabkan penumpukan stok barang yang tidak terjual, hal ini juga mempengaruhi keuntungan yang didapatkan. Oleh karena itu diperlukan adanya strategi yang tepat untuk mengetahui kebutuhan dan permintaan pelanggan pada suatu produk tertentu. Salah satu bentuk strategi bisnis adalah dengan melakukan pengelompokan produk yang memiliki karakter berbeda. Dengan melakukan pengelompokan produk dapat memudahkan pemilik bisnis mengontrol jumlah stok, meminimalisir kekurangan maupun kelebihan stok, dan pemenuhan permintaan konsumen dapat disiapkan tepat waktu. Untuk mendapatkan data produk yang banyak

diminati dan memiliki potensi memaksimalkan keuntungan tersebut bisa dengan melakukan analisis dan pengelompokan pada data transaksi produk.

Berdasarkan latar belakang diatas, penelitian ini mengusulkan segmentasi data produk berdasarkan model analisis RFM dan metode K-means *clustering*. RFM banyak digunakan dalam bidang pemasaran[2]. Dengan menggunakan model analisis RFM, seorang pengambil keputusan dapat mengidentifikasi *clustering* yang berharga dan mengembangkan strategi pemasaran dengan lebih efektif [3]. Atribut R mengacu pada jarak antara waktu terakhir transaksi dengan waktu analisis. Semakin pendek jaraknya, maka semakin besar nilai R. Selanjutnya, atribut F mewakili *recency*, yang mengacu pada jumlah transaksi dalam periode tertentu. Semakin banyak frekuensinya, semakin besar F. atribut M mewakili *recency*, yang mengacu pada jumlah uang konsumsi dalam periode tertentu. Banyaknya *recency*, maka nilai M lebih besar[4].

Dalam beberapa penelitian sebelumnya, Yaqin et al [5] menggunakan analisis RFM dan metode Fuzzy Sugeno untuk menentukan indeks produk pada permainan hayday, dalam penelitian ini analisis RFM digunakan sebagai atribut dalam mengukur nilai produk. Sedangkan fuzzy sugeno digunakan untuk proses penentuan nilai indeks produk. Dari penelitian ini didapatkan data masing – masing indeks produk pada permainan hayday yang bermanfaat untuk membantu memaksimalkan keuntungan dari penjualan produk. Dalam penelitian lain, Kurniati [6] menggunakan Model Analisis RFM untuk menentukan nilai indeks fitur ATM yang digunakan dalam segmentasi fitur ATM di PT BANK SYARIAH. Hasil dari penelitian ini didapatkan 3 *cluster* fitur ATM dengan kriteria yang berbeda. Berdasarkan beberapa penelitian yang dilakukan sebelumnya, maka timbul gagasan bahwa konsep RFM juga dapat diterapkan untuk segmentasi produk berdasarkan histori data transaksi penjualan.

Kemudian untuk proses pengelompokan pada penelitian ini menggunakan teknik *clustering*. *Clustering* adalah sebuah proses data mining yang digunakan untuk membuat kelompok dari kumpulan data berdasarkan kesamaan antar data tersebut [7]. Salah satu metode yang banyak digunakan dalam proses *clustering* adalah K-Means *clustering*[8]. K-means pertama kali ditemukan oleh Lloyd pada tahun 1957. Kemudian pada tahun 1965 Forgey juga menerbitkan metode yang sama, yang disebut Lloyd-Forgey. Metode K-Means mengelompokkan data yang ada ke dalam *cluster*. Di mana data yang ada dalam satu *cluster* memiliki karakteristik yang sama satu sama lain tetapi memiliki karakteristik yang berbeda dengan data di *cluster* lain[7].

Dalam penelitian sebelumnya, Event[1], menggunakan metode k-means *clustering* untuk studi kasus yang sama. Penelitian ini mendapatkan hasil berupa 3 *cluster* data produk. Dalam penelitian ini, data uji di kelompokkan menggunakan metode k-means *clustering* tanpa menentukan atribut yang digunakan sebagai parameter dalam proses *clustering*, penentuan jumlah *cluster* telah ditetapkan oleh peneliti sebelum proses *clustering*. Pada penelitian ini tidak dilakukan pengujian hasil *clustering* sehingga tidak diketahui apakah metode yang di gunakan efektif dalam menyelesaikan masalah segmentasi produk dan tidak diketahui bagaimana kualitas hasil *clustering* dengan metode yang telah digunakan.

Dalam penelitian yang akan dilakukan, proses *clustering* dilakukan 5 kali dengan jumlah *cluster* 2, 3, 4, 5 dan 6. Proses ini dilakukan dengan tujuan untuk mendapatkan jumlah *cluster* yang paling optimal dalam pengelompokan data produk yang ada. Hasil *clustering* di evaluasi menggunakan metode *Silhouette Coefficient*. *Silhouette Coefficient* adalah metode yang digunakan untuk mengukur kualitas serta kekuatan *cluster* dan seberapa baik suatu data ditempatkan pada suatu *cluster*[9]. Adapun *Silhouette coefficient* terdapat pada angka antara nilai -1 sampai 1 dimana nilai *Silhouette coefficient* semakin mendekati nilai 1, maka semakin bagus pengelompokan objek-objek kedalam sebuah *cluster* dan sebaliknya

jika *Silhouette coefficient* sudah mendekati angka -1, maka akan makin buruk metode pengelompokan datanya pada *cluster* tersebut[10].

METODE

Gambar 1. Metodologi Penelitian

Pada Gambar 1, dijelaskan tentang langkah-langkah penelitian. Dimulai dengan pengumpulan data transaksi penjualan dilakukan dengan memeriksa data dan membersihkan data data yang tidak valid serta data yang tidak diperlukan. Kemudian analisis RFM digunakan untuk menentukan nilai *recency*, *recency*, dan *recency* pada masing – masing produk. Selanjutnya nilai RFM masing – masing produk dinormalisasi menggunakan normalisasi standar. Untuk proses selanjutnya adalah *clustering* data penjualan produk menggunakan k-means *clustering*, dari hasil clastering didapatkan data segmentasi produk , hasil *clustering* tersebut diuji menggunakan rumus *Silhoutte Coefficient*.

Pengumpulan Data

Data yang dipakai dalam penelitian ini adalah data transaksi produk. Pengumpulan data dilakukan dengan melakukan analisis pada data transaksi produk dengan menghapus data – data yang tidak diperlukan dalam penelitian. Data transaksi yang digunakan pada penelitian ini berjumlah 88 data transaksi produk dengan periode transaksi dari tanggal 01 Agustus 2018 sampai tanggal 31 Agustus 2018.

Analisis RFM

Proses selanjutnya adalah analisis RFM. Analisis RFM pada penelitian ini digunakan untuk menentukan nilai masing – masing produk yang nantinya akan digunakan sebagai atribut dalam proses *clustering*. Adapun penentuan nilai dari atribut RFM [11], nilai *recency*, dilakukan dengan perhitungan selisih antara waktu terakhir transaksi produk dengan waktu dilakukannya penelitian. Dalam penentuan nilai *recency*, dibutuhkan atribut yang merepresentasikan berapa kali produk tersebut terlibat transaksi, untuk mendapatkan nilai *recency* dapat dilihat dari berapa banyak id produk yang sama muncul dalam data transaksi pada periode yang telah ditentukan. Untuk penentuan nilai *recency* adalah dengan menghitung total harga masing-masing produk yang terjual dalam data transaksi pada masa periode yang telah ditentukan. Adapun masa periode dalam penelitian ini adalah tanggal 12 September 2018.

Normalisasi Nilai RFM

Tahap selanjutnya dalam penelitian ini adalah normalisasi nilai RFM. Normalisasi nilai RFM ini dilakukan agar skala antara nilai data *recency*, *recency*, dan *recency* tidak berbeda jauh. Adapun untuk normalisasi dilakukan dengan membagikan nilai setiap data RFM dengan nilai RFM yang terbesar pada kolom RFM tersebut. Dari data Nilai RFM didapatkan nilai maksimum koefisien normalisasi yang terdiri dari R bernilai 42, F bernilai 8, dan M bernilai 2.500.000. Setelah diperoleh nilai tertinggi dari setiap kolom RFM, dilakukan normalisasi dengan menggunakan Persamaan (1) :

$$\text{Nilai (RFM)} = \frac{\text{Nilai RFM data ke } i}{\text{Nilai RFM max}} \quad (1)$$

Persamaan (2) adalah contoh perhitungan untuk normalisasi nilai R pada data ke 1:

$$\begin{aligned} R &= \frac{\text{Nilai R data ke } i}{\text{Nilai R max}} \\ R1 &= \frac{34}{42} \\ R1 &= 0,8095 \end{aligned} \quad (2)$$

Clustering Dengan Algoritma K-Means Clustering

Algoritma K-Means dalam penelitian ini digunakan dalam proses klasterisasi data transaksi produk berdasarkan nilai RFM. Pada penelitian ini, dilakukan 5 kali *clustering* mulai dari 2 *cluster* hingga 6 *cluster*. Berdasarkan penelitian sebelumnya, berikut adalah langkah – langkah yang dilakukan dalam proses *clustering* [12]:

1. Menentukan titik centroid. Titik centroid ini ditentukan secara random dari data Nilai RFM yang ada. Titik centroid dari masing – masing pengujian dicantumkan pada tabel 2 :

Tabel 2. Nilai Centroid awal Cluster K-Means

JUMLAH CLUSTER	CENTROID	R	F	M
2	C1	0,5710	0,3750	0,1740
	C2	0,3330	0,1250	0,0920
3	C1	0,7619	0,5000	0,3520
	C2	0,6429	0,1250	0,0360
	C3	0,3333	0,6250	0,1188
4	C1	1,0000	0,1250	0,0254

	C2	0,2857	0,2500	0,0900
	C3	0,4286	0,1250	0,0900
	C4	0,9762	0,1250	0,0186
5	C1	1,0000	0,1250	0,0294
	C2	0,4524	0,6250	0,5400
	C3	0,4286	0,1250	0,1500
	C4	0,8571	0,1250	0,0486
	C5	0,2857	0,1250	0,2660
6	C1	0,8095	0,2500	0,2800
	C2	0,7381	0,1250	0,0186
	C3	0,4524	0,1250	0,0060
	C4	0,4762	0,2500	0,0450
	C5	0,2857	0,2500	0,1040
	C6	0,5000	0,6250	0,2560

2. Hitung jarak dari setiap data ke pusat kluster. Untuk menghitung jarak dari setiap data ke pusat kluster dapat digunakan rumus euclidean pada Persamaan (3).

$$D(i, k) = \sqrt{(\sum_{j=1}^m (i_j - k_j)^2)} \quad (3)$$

3. Tahap selanjutnya yaitu menentukan pusat kluster baru dengan cara menghitung nilai rata – rata dari keseluruhan nilai masing - masing kriteria pada kluster yang sama.
4. Ulangi perhitungan jarak setiap data ke pusat kluster hingga posisi data sudah tidak berubah.

Pengujian Hasil *Clustering* Menggunakan Metode *Silhouette Coefficient*

Setelah dilakukan proses *clustering* pada data transaksi penjualan produk, maka dilakukan uji coba nilai silhouette dari hasil *clustering* masing – masing pengujian untuk mengetahui validasi dari hasil *clustering* dan mengidentifikasi kualitas dari *cluster* yang dihasilkan. Dimulai dengan menentukan rata rata jarak produk dengan semua produk lain dalam satu *cluster* menggunakan Persamaan (4):

$$a(i) = \frac{1}{[A]-1} \sum_{j \in A, j \neq i} d(i, j) \quad (4)$$

Dimana nilai A merupakan jumlah anggota dalam satu *cluster*. Selanjutnya adalah dengan menghitung rata – rata jarak produk dengan semua produk lain yang berada pada *cluster* lain, selanjutnya ambil nilai paling minimum dengan Persamaan (5):

$$d(i, c) = \frac{1}{[A]} \sum_{j \in C} d(i, j) \quad (5)$$

Nilai $d(i, c)$ yang dihasilkan akan memiliki jumlah nilai yang sama dengan jumlah kluster pada masing – masing pengujian. Dari nilai $d(i, c)$ akan diambil nilai minimum untuk mendapatkan nilai $b(i)$. Hitung nilai silhouette coefficient dengan Persamaan (6) :

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

Nilai $\max(a(i), b(i))$ merupakan nilai tertinggi maksimum antara $(a(i))$ dan suatu data $(b(i))$. Setelah didapatkan nilai $s(i)$ masing masing pengujian, selanjutnya diambil nilai rata – rata dari masing masing pengujian, nilai ini digunakan sebagai nilai *Silhouette Coefficient* dari hasil *clustering* menggunakan metode *k-means clustering*.

HASIL DAN PEMBAHASAN

Analisis RFM

Nilai *recency*, dilakukan dengan perhitungan selisih antara waktu terakhir transaksi produk dengan waktu dilakukannya penelitian. Dalam penentuan nilai *recency*, dibutuhkan atribut yang merepresentasikan berapa kali produk tersebut terlibat transaksi, untuk mendapatkan nilai *recency* dapat dilihat dari berapa banyak id produk yang sama muncul dalam data transaksi pada periode yang telah ditentukan. Untuk penentuan nilai *recency* adalah dengan menghitung total harga masing-masing produk yang terjual dalam data transaksi pada masa periode yang telah ditentukan. Adapun masa periode dalam penelitian ini adalah tanggal 12 September 2018. Setelah dilakukan analisis RFM pada data transaksi produk, maka didapatkan 44 data produk dengan nilai RFM seperti yang ditampilkan pada Tabel 1.

Normalisasi Nilai RFM

Dari data Nilai RFM didapatkan nilai maksimum koefisien normalisasi yang terdiri dari R bernilai 42, F bernilai 8, dan M bernilai 2.500.000. Setelah diperoleh nilai tertinggi dari setiap kolom RFM, dilakukan normalisasi dengan menggunakan Persamaan (1). Hasil perhitungan selengkapnya terdapat pada Tabel 1:

Tabel 1. Tabel Nilai RFM

PRODUK		TANPA NORMALISASI			NORMALISASI		
No	ID Produk	R	F	M	R	F	M
1	P01	34	2	Rp700.000	0,8095	0,2500	0,2800
2	P02	42	1	Rp73.500	1,0000	0,1250	0,0294
3	P03	41	1	Rp56.500	0,9762	0,1250	0,0226
4	P04	42	1	Rp63.500	1,0000	0,1250	0,0254
5	P05	12	1	Rp93.000	0,2857	0,1250	0,0372
6	P06	32	4	Rp880.000	0,7619	0,5000	0,3520
7	P07	32	2	Rp240.000	0,7619	0,2500	0,0960
8	P08	31	1	Rp46.500	0,7381	0,1250	0,0186
9	P09	19	5	Rp1.350.000	0,4524	0,6250	0,5400
10	P10	24	3	Rp435.000	0,5714	0,3750	0,1740
11	P11	12	2	Rp225.000	0,2857	0,2500	0,0900
12	P12	21	2	Rp500.000	0,5000	0,2500	0,2000
13	P13	25	2	Rp400.000	0,5952	0,2500	0,1600
14	P14	24	5	Rp1.050.000	0,5714	0,6250	0,4200
15	P15	24	1	Rp170.000	0,5714	0,1250	0,0680
16	P16	12	8	Rp2.500.000	0,2857	1,0000	1,0000
17	P17	38	1	Rp20.000	0,9048	0,1250	0,0080
18	P18	19	1	Rp15.000	0,4524	0,1250	0,0060

19	P19	22	1	Rp210.000	0,5238	0,1250	0,0840
20	P20	18	1	Rp375.000	0,4286	0,1250	0,1500
21	P21	30	1	Rp840.000	0,7143	0,1250	0,3360
22	P22	27	1	Rp10.000	0,6429	0,1250	0,0040
23	P23	38	1	Rp270.000	0,9048	0,1250	0,1080
24	P24	27	1	Rp90.000	0,6429	0,1250	0,0360
25	P25	12	5	Rp600.000	0,2857	0,6250	0,2400
26	P26	20	2	Rp112.500	0,4762	0,2500	0,0450
27	P27	36	1	Rp121.500	0,8571	0,1250	0,0486
28	P28	18	1	Rp225.000	0,4286	0,1250	0,0900
29	P29	39	1	Rp111.500	0,9286	0,1250	0,0446
30	P30	40	1	Rp130.000	0,9524	0,1250	0,0520
31	P31	33	1	Rp145.000	0,7857	0,1250	0,0580
32	P32	14	1	Rp230.000	0,3333	0,1250	0,0920
33	P33	12	2	Rp260.000	0,2857	0,2500	0,1040
34	P34	12	3	Rp1.625.000	0,2857	0,3750	0,6500
35	P35	12	1	Rp665.000	0,2857	0,1250	0,2660
36	P36	34	1	Rp86.500	0,8095	0,1250	0,0346
37	P37	27	1	Rp60.000	0,6429	0,1250	0,0240
38	P38	14	5	Rp297.000	0,3333	0,6250	0,1188
39	P39	12	3	Rp750.000	0,2857	0,3750	0,3000
40	P40	19	1	Rp200.000	0,4524	0,1250	0,0800
41	P41	41	1	Rp46.500	0,9762	0,1250	0,0186
42	P42	22	2	Rp50.000	0,5238	0,2500	0,0200
43	P43	24	1	Rp78.000	0,5714	0,1250	0,0312
44	P44	21	5	640000	0,5000	0,6250	0,2560

Clustering Dengan Algoritma K-Means Clustering

Algoritma K-Means dalam penelitian ini digunakan dalam proses klasterisasi data transaksi produk berdasarkan nilai RFM. Pada penelitian ini, dilakukan 5 kali *clustering* mulai dari 2 *cluster* hingga 6 *cluster*:

1. Menentukan titik centroid. Titik centroid ini ditentukan secara random dari data Nilai RFM yang ada. Titik centroid dari masing – masing pengujian dicantumkan pada Tabel 2 :

Tabel 2. Nilai Centroid awal *Cluster* K-Means

JUMLAH CLUSTER	CENTROID	R	F	M
2	C1	0,5710	0,3750	0,1740
	C2	0,3330	0,1250	0,0920
3	C1	0,7619	0,5000	0,3520
	C2	0,6429	0,1250	0,0360
	C3	0,3333	0,6250	0,1188
4	C1	1,0000	0,1250	0,0254
	C2	0,2857	0,2500	0,0900

	C3	0,4286	0,1250	0,0900
	C4	0,9762	0,1250	0,0186
5	C1	1,0000	0,1250	0,0294
	C2	0,4524	0,6250	0,5400
	C3	0,4286	0,1250	0,1500
	C4	0,8571	0,1250	0,0486
	C5	0,2857	0,1250	0,2660
6	C1	0,8095	0,2500	0,2800
	C2	0,7381	0,1250	0,0186
	C3	0,4524	0,1250	0,0060
	C4	0,4762	0,2500	0,0450
	C5	0,2857	0,2500	0,1040
	C6	0,5000	0,6250	0,2560

2. Hitung jarak dari setiap data ke pusat kluster. Hasil perhitungan *clustering* 1 menggunakan Persamaan (3) terdapat pada Tabel 3:

Tabel 3. Jarak Tiap Data Terhadap Setiap Centroid *Cluster* (*clustering* 1)

Nama Produk	C1	C2	Cluster
P01	0,289	0,527	C1
P02	0,517	0,670	C1
P03	0,499	0,647	C1
P04	0,518	0,670	C1
P05	0,404	0,073	C2
P06	0,289	0,626	C1
P07	0,241	0,446	C1
P08	0,338	0,411	C1
P09	0,459	0,682	C1
P10	0,000	0,355	C1
P11	0,323	0,134	C2
P12	0,146	0,235	C1
P13	0,128	0,298	C1
P14	0,351	0,644	C1
P15	0,272	0,239	C2
P16	1,074	1,262	C1
P17	0,449	0,578	C1
P18	0,324	0,147	C2
P19	0,270	0,191	C2
P20	0,289	0,112	C2
P21	0,330	0,452	C1
P22	0,311	0,322	C1
P23	0,422	0,572	C1
P24	0,294	0,315	C1
P25	0,385	0,524	C1

P26	0,203	0,196	C2
P27	0,400	0,526	C1
P28	0,300	0,095	C2
P29	0,455	0,597	C1
P30	0,472	0,620	C1
P31	0,349	0,454	C1
P32	0,355	0,000	C2
P33	0,320	0,134	C2
P34	0,555	0,613	C1
P35	0,391	0,180	C2
P36	0,372	0,480	C1
P37	0,300	0,317	C1
P38	0,350	0,501	C1
P39	0,312	0,329	C1
P40	0,292	0,120	C2
P41	0,500	0,647	C1
P42	0,204	0,239	C1
P43	0,288	0,246	C2
P44	0,273	0,552	C1

Kemudian untuk *clustering* 2 sampai *clustering* 5 juga dilakukan dengan proses yang sama pada data uji yang telah ditentukan. Pada hasil *clustering* tersebut data yang memiliki jarak paling minimum dari suatu kluster akan menjadi anggota dari kluster tersebut.

3. Tahap selanjutnya yaitu menentukan pusat kluster baru dengan cara menghitung nilai rata – rata dari keseluruhan nilai masing - masing kriteria pada kluster yang sama. Tabel 4 adalah tabel hasil perhitungan centroid dari *clustering* 1:

Tabel 4. Nilai Centroid Baru (*clustering* 1)

	R	F	M
C1	0,677	0,290	0,183
C2	0,414	0,154	0,088

Untuk penentuan titik centroid pada *clustering* 2, *clustering* 3, *clustering* 4, dan *clustering* 5 juga dilakukan dengan proses yang sama

4. Ulangi perhitungan jarak setiap data ke pusat kluster hingga posisi data sudah tidak berubah. Pada penelitian ini, Proses ini juga dilakukan pada *clustering* 3, 4, dan 5. Pada *clustering* 1, proses *clustering* dan perhitungan jarak berhenti pada iterasi ke 6 dengan anggota C1 = 20 data dan anggota C2 = 22 data. Pada tabel 5 berikut adalah hasil dari proses *clustering* pada pengujian pertama:

Tabel 5. Hasil proses *clustering* (*clustering* 1)

Nama Produk	C1	C2	Cluster
P01	0,2407	0,3996	C1
P02	0,1946	0,6469	C1
P03	0,1730	0,6280	C1

P04	0,1954	0,6481	C1
P05	0,5248	0,3219	C2
P06	0,4635	0,3918	C2
P07	0,1255	0,3742	C1
P08	0,0873	0,4359	C1
P09	0,7673	0,4209	C2
P10	0,3529	0,1596	C2
P11	0,5362	0,2172	C2
P12	0,3551	0,1236	C2
P13	0,2592	0,2063	C2
P14	0,6472	0,3708	C2
P15	0,2384	0,3083	C1
P16	1,3742	1,0226	C2
P17	0,1128	0,5717	C1
P18	0,3626	0,3130	C2
P19	0,2865	0,2801	C2
P20	0,3901	0,2323	C2
P21	0,2855	0,3787	C1
P22	0,1787	0,3817	C1
P23	0,1044	0,5411	C1
P24	0,1700	0,3639	C1
P25	0,7361	0,3139	C2
P26	0,3525	0,2121	C2
P27	0,0526	0,5169	C1
P28	0,3818	0,2587	C2
P29	0,1218	0,5792	C1
P30	0,1442	0,5978	C1
P31	0,0284	0,4552	C1
P32	0,4770	0,2733	C2
P33	0,5370	0,2085	C2
P34	0,8188	0,4452	C2
P35	0,5604	0,2619	C2
P36	0,0349	0,4832	C1
P37	0,1726	0,3703	C1
P38	0,6834	0,3153	C2
P39	0,6205	0,1595	C2
P40	0,3576	0,2656	C2
P41	0,1741	0,6293	C1
P42	0,3107	0,2489	C2
P43	0,2411	0,3289	C1
P44	0,6075	0,2925	C2

Pada *clustering* 2, proses *clustering* dan perhitungan jarak berhenti pada iterasi ke 7 dengan anggota C1 = 8 data, anggota C2 = 18 data, dan anggota C3 = 18 data. Pada *clustering* 3, proses *clustering* dan perhitungan jarak berhenti pada iterasi ke 4 dengan anggota C1 = 9 data, anggota C2 = 9 data, anggota C3 = 17, dan anggota C4 = 9. Pada *clustering* 4, proses *clustering* dan perhitungan jarak berhenti pada iterasi ke 5 dengan anggota C1 = 9 data, anggota C2 = 8 data, anggota C3 = 12 data, anggota C4 = 9, dan anggota C5 = 6. Pada *clustering* 5, proses *clustering* dan perhitungan jarak berhenti pada iterasi ke 2 dengan anggota C1 = 3 data, anggota C2 = 13 data, dan anggota C3 = 10 data, anggota C4 = 5 data, anggota C5 = 6 data, dan anggota C6 = 7 data.

Pengujian Hasil *Clustering* Menggunakan Metode *Silhouette Coefficient*

Setelah dilakukan proses *clustering* pada data transaksi penjualan produk, maka dilakukan uji coba nilai *silhouette* dari hasil *clustering* masing – masing pengujian untuk mengetahui validasi dari hasil *clustering* dan mengidentifikasi kualitas dari *cluster* yang dihasilkan. Dimulai dengan menentukan rata rata jarak produk dengan semua produk lain dalam satu *cluster* menggunakan Persamaan (4). Jika diimplementasikan dalam data ke 1 pada *clustering* ke 1 adalah sebagai berikut:

$$= \frac{\sqrt{(1,0000 - 0,8095)^2} + (0,1250 - 0,2500)^2 + (0,2994 - 0,2800)^2 + \dots + \sqrt{(0,5714 - 0,8095)^2} + (0,1250 - 0,2500)^2 + (0,0312 - 0,2800)^2}{19}$$

$$= 0,2961$$

Selanjutnya adalah dengan menghitung rata – rata jarak produk dengan semua produk lain yang berada pada *cluster* lain, selanjutnya ambil nilai paling minimum dengan Persamaan (5). Jika diimplementasikan dalam data ke 1 pada *clustering* ke 1 adalah sebagai berikut:

$$= \frac{\sqrt{(0,8095 - 0,2857)^2} + (0,2500 - 0,1250)^2 + (0,2800 - 0,0372)^2 + \dots + \sqrt{(0,5714 - 0,2857)^2} + (0,1250 - 0,1250)^2 + (0,0312 - 0,0372)^2}{20}$$

$$= 0,5141$$

Hitung nilai *silhouette coefficient* dengan Persamaan (6). Jika diimplementasikan dalam data ke 1 pada *clustering* ke 1 adalah sebagai berikut:

$$= \frac{0,5984 - 0,2961}{0,5984}$$

$$= 0,5053$$

Selanjutnya diambil nilai rata – rata dari masing masing pengujian, nilai ini digunakan sebagai nilai *Silhouette Coefficient* dari hasil *clustering* menggunakan metode k-means *clustering*. Dari Persamaan (6) didapatkan nilai *Silhouette Coefficient* global yang ditampilkan pada Tabel 6.

Tabel 6. Perhitungan *Silhouette Coefficient* (pengujian 1)

JUMLAH <i>CLUSTER</i>	GLOBAL <i>SILHOUTTE</i>
2	0,2890
3	0,4314
4	0,3567
5	0,3433
6	0,1962

Semakin nilai *Silhouette coefficient* mendekati nilai 1, maka semakin baik pengelompokan data dalam satu *cluster*. Sebaliknya jika *Silhouette coefficient* mendekati nilai -1, maka semakin buruk pengelompokan data didalam satu *cluster*. Berdasarkan uji coba yang telah dilakukan diketahui bahwa nilai *Silhouette Coefficient* global pada 2 *cluster* adalah 0,2890, pada 3 *cluster* 0,4314, pada 4 *cluster* 0,4567, pada 5 *cluster* adalah 0,3433, dan pada 6 *cluster* adalah 0,1962. Adapun hasil dari pengujian *Silhouette Coefficient* pada masing – masing proses *clustering* direpresentasikan pada Gambar 2.

Gambar 2. Grafik nilai *Silhouette Coefficient*

Berdasarkan grafik *Silhouette Coefficient* pada Gambar 3. Diketahui bahwa diantara 5 kali proses *clustering* yang dilakukan, nilai *Silhouette Coefficient* global tertinggi ada pada *clustering* ke 2 dengan jumlah *cluster* sebanyak 3. Dari 44 data transaksi produk, setelah melalui proses *clustering* didapatkan bahwa pada C1 berisi 8 anggota data produk, pada C2 berisi 18 anggota data produk, dan pada C3 berisi 18 anggota data produk. Untuk mengetahui karakteristik produk pada masing – masing *cluster*, maka dihitung nilai rata rata *recency*, *recency*, dan *recency* pada masing masing *cluster*. Tabel 7 adalah hasil perhitungan rata – rata nilai RFM pada Pengujian 2 :

Tabel 7. Perhitungan rata – rata RFM masing – masing *cluster*

	C1	C2	C3
R	0,4345	0,8360	0,4365
F	0,6250	0,1389	0,1944
M	0,4471	0,0691	0,1110

Berdasarkan pengertian RFM dimana nilai *recency* merupakan jarak antara tanggal terakhir transaksi dengan tanggal periode analisis maka semakin kecil nilai *recency* semakin bagus karena dekatnya jarak produk tersebut terakhir dibeli pelanggan. Sedangkan semakin besar nilai *recency* menandakan semakin baik *cluster* tersebut karena itu berarti semakin sering produk dalam *cluster* tersebut dibeli oleh pelanggan. Untuk nilai *recency*, semakin besar nilai *recency* menandakan semakin baik *cluster* tersebut karena itu berarti produk dalam *cluster* tersebut adalah produk yang berpotensi untuk dimaksimalkan keuntungannya.

Gambar 3. Grafik hasil *clustering*

Berdasarkan Gambar 3, dapat diketahui bahwa pada C1 berupa kelompok produk dengan nilai *recency*, *recency*, dan *recency* paling baik. Pada C2 berisi kelompok produk yang memiliki nilai *recency*, *recency*, dan nilai *recency* paling buruk. Pada C3 berisi kelompok produk yang memiliki nilai *recency*, *recency*, dan *recency* yang cukup baik.

KESIMPULAN

Tujuan utama dalam penelitian ini adalah untuk segmentasi produk dengan mengkombinasikan model analisis RFM dan *clustering* menggunakan metode K-Means *clustering*. Dalam pengujian ini, dilakukan 5 kali proses *clustering* dengan jumlah *cluster* yang berbeda. Hasil masing – masing *clustering* di evaluasi menggunakan metode *Silhouette Coefficient* untuk mengetahui jumlah *cluster* yang paling optimal dalam pengelompokan data yang ada. Dari hasil evaluasi didapatkan nilai *silhouette* tertinggi adalah 0,4314 pada *clustering* ke 2 dengan jumlah 3 *cluster*. Dari penelitian ini didapatkan 3 kelompok produk dengan karakter yang berbeda. C1 berupa kelompok produk dengan nilai *recency*, *recency*, dan *recency* paling baik. Pada C2 berisi kelompok produk yang memiliki nilai *recency* dan *recency* yang baik namun nilai *recency* paling buruk. Pada C3 berisi kelompok produk yang memiliki nilai *recency*, *recency*, dan *recency* yang cukup baik. Dengan demikian dapat diketahui bahwa C1 merupakan kelompok produk yang banyak diminati serta memiliki potensi untuk memaksimalkan keuntungan pemilik bisnis, sehingga bisa ditambahkan stok persediaan untuk produk – produk ini. Kemudian pada C2 berisi kelompok produk yang tidak diminati, tidak memiliki potensi untuk memaksimalkan keuntungan, sehingga tidak perlu terlalu banyak stok yang disediakan agar tidak terjadi penumpukan stok. Pada C3 berisi

kelompok produk dengan peminat dan potensi keuntungan menengah, untuk produk ini bisa dilakukan persediaan stock yang sama seperti sebelumnya, karena penjualan pada produk ini ada di tingkat tengah.

REFERENSI

- [1] Event and I. Utnasari, "Analisis *Clustering* dengan K-Means untuk Pengelompokkan Penjualan Produk Pada Hotel Newton," *Comput. Sci. Ind. Eng.*, vol. 04, no. Vol 4 No 4 (2021): Comasie, pp. 1–8, 2021, [Online]. Available: <http://ejournal.upbatam.ac.id/index.php/comasiejournal/article/view/3430>.
- [2] R. C. Blattberg, B.-D. Kim, and S. A. Neslin, *RFM Analysis*, no. M. Springer, New York, NY, 2008.
- [3] J. Wei, S. Lin, and H. Wu, "A review of the application of RFM model," *African J. Bus. Manag.*, vol. 4, no. 19, pp. 4199–4206, 2010, doi: <https://doi.org/10.5897/AJBM.9000026>.
- [4] C. H. Cheng and Y. S. Chen, "Classifying the segmentation of *clustering* value via RFM model and RS theory," *Expert Syst. Appl.*, vol. 36, no. 3 PART 1, pp. 4176–4184, 2009, doi: [10.1016/j.eswa.2008.04.003](https://doi.org/10.1016/j.eswa.2008.04.003).
- [5] M. A. Yaqin, S. Naja, S. K. Al-Azhar, and K. Mahbullah, "Penerapan Metode Fuzzy Sugeno pada Analisis RFM untuk Menentukan Indeks Produk pada Permainan Hayday," in *Prosiding SENIATI*, 2019, vol. 5, no. Prosiding SENIATI 2019 (BOOK-1), pp. 50–56, [Online]. Available: [internal-pdf://125.241.156.57/Analisis RFM untuk Menentukan Indeks Produk.pdf](https://internal-pdf://125.241.156.57/Analisis%20RFM%20untuk%20Menentukan%20Indeks%20Produk.pdf).
- [6] E. A. Kurniati, R. Hendrawan, and R. A. Vinarti, "Visualisasi Segmentasi Fitur ATM Menggunakan Analisis RFM dan Algoritma K-Means untuk Mendukung Evaluasi ATM di PT. BANK," *J. Tek. POMITS*, vol. 1, no. 1, pp. 1–12, 2014, [Online]. Available: <https://repository.its.ac.id/352/1/5211100097-paperpdf.pdf>.
- [7] T. Widiyaningtyas, M. I. W. Prabowo, and M. A. M. Pratama, "Implementation of k-means *clustering* method to distribution of high school teachers," *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, vol. 2017-Decem, no. September, pp. 19–21, 2017, doi: [10.1109/EECSI.2017.8239083](https://doi.org/10.1109/EECSI.2017.8239083).
- [8] X. Wu *et al.*, *Top 10 algorithms in data mining*, vol. 14, no. 1. 2008.
- [9] F. G. Febrinanto, C. Dewi, and A. T. Wiratno, "Implementasi Algoritme K-Means Sebagai Metode Segmentasi Citra Dalam Identifikasi Penyakit Daun Jeruk," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 11, pp. 5375–5383, 2018.
- [10] I. M. Dwi Putra Asana, I. M. Oka Widyantara, N. M. A. E. . Wirastuti, and I. B. P. Adnyana, "Metode Contrast Stretching untuk Perbaikan Kualitas Citra pada Proses Segmentasi Video," *Maj. Ilm. Teknol. Elektro*, vol. 16, no. 2, p. 1, 2017, doi: [10.24843/mite.2017.v16i02p01](https://doi.org/10.24843/mite.2017.v16i02p01).
- [11] S. S. Prasetyo, M. Mustafid, and A. R. Hakim, "Penerapan Fuzzy C-Means Kluster Untuk Segmentasi Pelanggan E-Commerce Dengan Metode *Recency Recency Recency* (Rfm)," *J. Gaussian*, vol. 9, no. 4, pp. 421–433, 2020, doi: [10.14710/j.gauss.v9i4.29445](https://doi.org/10.14710/j.gauss.v9i4.29445).
- [12] N. Shi, X. Liu, and Y. Guan, "Research on k-means *clustering* algorithm: An improved k-means *clustering* algorithm," *3rd Int. Symp. Intell. Inf. Technol. Secur. Informatics, IITSI 2010*, pp. 63–67, 2010, doi: [10.1109/IITSI.2010.74](https://doi.org/10.1109/IITSI.2010.74).