

# IDENTIFIKASI KARAKTER PELANGGAN RETAIL BERPOTENSI *CHURN* DENGAN METODA *RECENCY*, *FREQUENCY* DAN *MONETARY*

Ni Wayan Wardani<sup>1)</sup> Eddy Hartono<sup>2)</sup>

Program Studi Teknik Informatika<sup>1)2)</sup>

STMIK STIKOM Indonesia, Denpasar, Bali<sup>1)2)</sup>

niwayan.wardani@stiki-indonesia.ac.id<sup>1)</sup> eddy\_h99@yahoo.com<sup>2)</sup>

## ABSTRACT

*Analyzing the character of the customer is very important for every company, especially retail companies, because of the increasingly fierce competition of retail companies where retail companies have mushroomed right up to the middle of residents' and remote villages. The character of customers varies greatly, from customers who have high loyalty to the character of potential customers who churn. Churn is a term used for customers who have stopped purchasing. By implementing the concept of CRM (Customer Relationship Management), companies can identify the character of potential churn customers. Knowing the character of customers can determine the company's policy to treat each of its customers and determine appropriate and effective marketing strategies. One reliable method that can be used to identify customer characters is the RFM method (recency, frequency, and monetary). From the segmentation process it can be seen that the customer class owned by UD. Mawar Sari namely the Dormant, Everyday, Golden and Superstar classes. Trial grouping of customer classes with K-Means Clustering on several clusters, namely 2 clusters up to 5 clusters, after being tested with the Davies-Boundin Index, there are 5 clusters as the best customer grouping. UD Mawar Sari customer class with potential churn is the that is in the second cluster and the third cluster, namely the customer class Dormant B, Dormant E, Dormant F and Everyday E.*

**Keywords :** *Churn, Retail, Recency, Frequency, Monetary*

## ABSTRAK

Menganalisa karakter pelanggan sangat penting bagi setiap perusahaan khususnya perusahaan retail, karena semakin ketatnya persaingan perusahaan retail dimana saat ini perusahaan retail telah menjamur sampai ke tengah – tengah perumahan penduduk dan pelosok desa . Karakter pelanggan sangat bervariasi, dari pelanggan yang memiliki loyalitas tinggi sampai karakter pelanggan yang berpotensi *churn*. *Churn* adalah istilah yang digunakan untuk pelanggan yang telah menghentikan pembelian. Dengan menerapkan konsep CRM (*Customer Relationship Management*), perusahaan dapat melakukan identifikasi karakter pelanggan berpotensi *churn*. Mengetahui karakter pelanggan dapat menentukan kebijakan perusahaan untuk memperlakukan setiap pelanggannya dan menentukan strategi pemasaran yang tepat dan efektif. Salah satu metoda handal yang dapat digunakan untuk identifikasi karakter pelanggan adalah metoda *RFM* (*recency, frequency, dan monetary*). Dari proses segmentasi tersebut dapat diketahui kelas pelanggan yang dimiliki oleh UD. Mawar Sari yaitu kelas *Dormant, Everyday, Golden* dan *Superstar*. Uji coba pengelompokkan kelas pelanggan dengan *K-Means Clustering* pada beberapa jumlah cluster yaitu 2 *cluster* sampai dengan 5 *cluster*, setelah diuji dengan *Davies-Boundin Index* mendapatkan jumlah 5 *cluster* sebagai pengelompokkan kelas pelanggan yang terbaik. Pelanggan UD. Mawar Sari yang berpotensi *churn* adalah kelas pelanggan yang berada di *cluster* ke-2 dan *cluster* ke-3 yaitu kelas pelanggan *Dormant B, Dormant E, Dormant F* dan *Everyday E*.

**Kata kunci :** *Churn, Retail, Recency, Frequency, Monetary*

## PENDAHULUAN

Pelanggan merupakan aset perusahaan yang harus dipertahankan, mengingat saat ini persaingan sangatlah ketat terutama di usaha retail. Tidak akan ada prospek bisnis tanpa adanya hubungan antara perusahaan dengan konsumen yang bersifat loyal. Hal ini adalah alasan mengapa perusahaan harus merencanakan dan menggunakan strategi yang cukup jelas dalam memperlakukan pelanggan. Perusahaan harus dapat mengenali dan mengetahui karakter pelanggan dengan baik. Karakter pelanggan menjadi dasar bagi perusahaan untuk mengambil berbagai kebijakan terkait berbagai hal seperti pemasaran, pembelian, promosi, dll.

Menurut Emmet C. Murphy dan Mark A. Murphy dalam buku *Leading On The Edge of Chaos* bahwa memperoleh pelanggan baru menghabiskan biaya lima kali lipat dari biaya untuk memuaskan dan mempertahankan pelanggan lama, sementara itu peningkatan sebanyak 2% dalam mempertahankan pelanggan (*customer retention*) punya dampak terhadap laba seperti memangkas biaya sebesar 10%. Oleh karena itu, dengan menerapkan metoda untuk mengetahui karakter pelanggan, perusahaan dapat melakukan identifikasi pelanggan berpotensi *churn* dan menerapkan strategi pemasaran yang tepat terhadap pelanggan – pelanggan lama dengan harapan dapat meningkatkan laba perusahaan.

Pada penelitian ini menggunakan metoda RFM (*recency, frequency, monetary*) untuk mengidentifikasi karakter pelanggan dan menghasilkan kelas pelanggan. Fitur *recency, frequency*, dan *monetary* di dapatkan dari data pelanggan dan data transaksi pelanggan. *Recency* adalah kapan transaksi terakhir dilakukan. *Frequency* adalah tingkat keseringan pelanggan melakukan transaksi, dan *monetary* adalah besarnya nilai transaksi yang dilakukan.

## TINJAUAN PUSTAKA

### *Customer Churn*

Menurut Yu dkk (2011), *churn* adalah saat dimana pelanggan memutuskan untuk berpindah dari satu layanan kepada layanan lain dengan jasa/produk yang sama. Permasalahan ini menjadi krusial yang harus dihadapi oleh berbagai perusahaan dalam era ini, disamping itu menurut Khakabi dkk (2010) untuk memperoleh pelanggan baru

memerlukan biaya 10x lebih mahal dari biaya untuk mempertahankan pelanggan yang ada. *Churn* dibagi 3 jenis, yaitu :

1. Aktif / disengaja : pelanggan memutuskan untuk menghentikan pembelian dan beralih ke perusahaan lain dengan alasan :
  - a. Tidak ada pemahaman tentang layanan skema
  - b. Tidak ada resolusi dari kesalahan
  - c. Biaya terlalu tinggi
  - d. Tidak ada imbalan loyalitas pelanggan
  - e. Masalah *privacy*
2. Rotasi / Insidental : pelanggan menghentikan pembelian tanpa bertujuan untuk berpindah ke perusahaan lain dengan alasan :
  - a. Masalah financial
  - b. Perubahan lokasi geografis pelanggan
  - c. Pelanggan sudah tidak memerlukan produk atau layanan yang disediakan perusahaan
3. Pasif : pihak perusahaan menghentikan kontraknya dengan alasan pelanggan tidak dapat melunasi tagihan.

### *Customer Relationship Management (CRM)*

Konsumen merupakan aset yang sangat penting bagi suatu perusahaan. Tidak akan ada prospek bisnis tanpa adanya hubungan antara perusahaan dengan konsumen yang bersifat loyal. Hal ini adalah alasan mengapa perusahaan harus merencanakan dan menggunakan strategi yang cukup jelas dalam memperlakukan konsumen. CRM telah berkembang dalam beberapa dekade belakangan ini untuk merefleksikan peranan utama dari konsumen untuk pengaturan strategi perusahaan. CRM meliputi seluruh ukuran untuk memahami konsumen dan proses untuk mengeksploitasi pengetahuan ini, untuk merancang dan mengimplementasikannya pada kegiatan marketing, produksi, dan rantai supply dari pemasok (*supplier*).

### Model RFM

Model RFM (*Recency, Frequency, Monetary*) adalah model berbasis perilaku digunakan untuk menganalisis perilaku pelanggan dan kemudian membuat prediksi berdasarkan perilaku database (Hughes, 2000). Model RFM ini merupakan metoda yang sudah lama dan populer untuk mengukur hubungan dengan pelanggan

(Schijns, 1996).

Model RFM diaplikasikan secara luas pada database pemasaran dan merupakan tool yang umum digunakan untuk membangun strategi pemasaran (Wei et al, 2010). RFM berdasarkan segmentasi pelanggan menghasilkan kemampuan antara 75% sampai 85% (Hughes, 2000). Berikut ini adalah penjelasan mengenai RFM menurut C.H.Cheng, Y.S Chen (2009) :

1. *Recency*, yaitu kapan transaksi terakhir dilakukan.  
Pelanggan yang baru saja melakukan transaksi pembelian dianggap lebih menjadi pelanggan aktif dibandingkan pelanggan yang telah lama melakukan transaksi (Wu and Cen,2000). Menurut Chen et al (2005), pada usaha retail, *recency* untuk kunjungan toko adalah dimensi yang paling penting dari model RFM. *Recency* mewakili jumlah hari antara akhir periode analisis dan tanggal transaksi terakhir. Untuk tujuan *model training*, *recency of churners* adalah jumlah hari antara tanggal diklasifikasikannya seorang pelanggan memiliki kecenderungan untuk *churn* dengan tanggal terakhir transaksi pembelian.
2. *Frequency*, yaitu tingkat keseringan pelanggan melakukan transaksi. Misalkan sekali transaksi tiap bulan atau 2 kali dalam setahun. *Frequency* adalah ukuran dari kekuatan hubungan antara pelanggan dan perusahaan. Pelanggan yang loyal didefinisikan adalah pelanggan yang lebih sering melakukan pembelian dibandingkan dengan pelanggan yang tidak loyal (Kamakura et al, 1991). Menurut Berry dan Linoff (2004), *frequency* adalah predictor penting dari *churn*. Untuk tujuan model training, *frequency of churners* adalah jumlah transaksi yang telah dilakukan dari tanggal pengamatan sampai dengan tanggal terakhir transaksi pembelian.
3. *Monetary*, besarnya nilai transaksi yang dilakukan.  
*Monetary value* adalah ukuran pengeluaran pelanggan pada perusahaan tertentu. Menurut Schmittlein dan Peterson (1994), nilai moneter pembelian masa lalu masing-masing pelanggan dapat menjadi prediktor penting dari perilaku masa depan. Untuk tujuan model training, *monetary of churners* adalah besarnya nilai transaksi pelanggan dari tanggal pengamatan

sampai dengan tanggal terakhir transaksi pembelian.

### **K-Means Clustering**

Metode ini termasuk teknik penyekatan (*partition*) yang membagi atau memisahkan objek ke- k daerah bagian yang terpisah. Pada *K-Means*, setiap objek harus masuk dalam kluster tertentu, tetapi dalam satu tahapan berikutnya objek akan berpindah ke kluster lain. Prosedur dalam algoritma *K-Means* dijelaskan pada tahap berikut :

1. Kelompokkan item dalam dataset ke dalam *K cluster*
2. Menentukan secara acak nilai *centroid*
3. Memasukkan setiap item dataset yang jaraknya paling dekat dengan nilai *centroid* ke dalam *centroid cluster* tersebut
4. Menghitung rata-rata nilai item dalam setiap *cluster* untuk dijadikan sebagai *centroid* yang baru
5. Melakukan pengulangan langkah 2 dan langkah 3 hingga nilai *centroid* sama dengan nilai rata – rata item dalam *cluster*

Perhitungan jarak antar titik dengan menggunakan *Euclidean distance* dapat dilakukan pada titik dalam satu dimensi, dua dimensi, ataupun 3 dimensi. Formula jarak antar dua titik dalam satu, dua dan tiga dimensi secara berurutan ditunjukkan pada formula 1,2,3 berikut ini :

$$\sqrt{(x + y)^2} = |x - y| \dots\dots\dots(1)$$

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} \dots\dots\dots(2)$$

$$d(p,q)=\sqrt{(p1-q1)^2 + (p2-q2)^2 + (p3-q3)^3} \dots(3)$$

### **Davies-Bouldin Index**

Metrik Davies-Bouldin Index (DBI) diperkenalkan oleh David L. Davies dan Donald W. Bouldin (1979) yang digunakan untuk mengevaluasi kluster. Validasi internal yang dilakukannya adalah seberapa baik clustering sudah dilakukan dengan menghitung kuantitas dan fitur turunan dari set data.

*Sum of square within cluster* (SSW) sebagai metrik kohesi dalam sebuah *cluster* ke – I diformulasikan oleh persamaan :

$$SSWi = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i)$$

Keterangan :

- $m_i$  : jumlah data yang berada dalam kluster ke  $- i$
- $X_j$  : nilai data ke-  $i$
- $C_i$  : *centroid* kluster ke- $i$

Sementara untuk menghitung jarak (*Euclidean*) antar *centroid* menggunakan persamaan :

$$SSB_{i,j} = d(c_i, c_j)$$

Keterangan :

- $C_i$  : *centroid* ke- $i$
- $C_j$  : *centroid* ke- $j$

Selanjutnya adalah menghitung ukuran rasio seberapa baik nilai perbandingan antara kluster ke- $i$  dan kluster ke- $j$  menggunakan formulasi persamaan :

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

Sifat – sifat yang dimiliki  $R_{i,j}$  sebagai berikut :

1.  $R_{i,j} \geq 0$
2.  $R_{i,j} = R_{j,i}$
3. Jika  $SSW_j \geq SSW_r$  dan  $SSB_{i,j} = SSB_{i,r}$  maka  $R_{i,j} > R_{i,r}$
4. Jika  $SSW_j = SSW_r$  dan  $SSB_{i,j} \leq SSB_{i,r}$  maka  $R_{i,j} > R_{i,r}$

Nilai *Davies-Bouldin Index* (DBI) didapatkan dari persamaan berikut :

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{i,j})$$

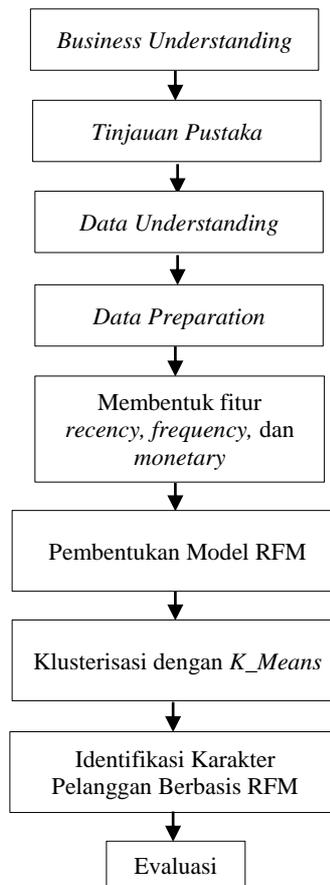
$K$  adalah jumlah kluster yang digunakan.

Dari syarat-syarat perhitungan yang didefinisikan di atas, dapat diamati bahwa semakin kecil nilai  $SSW$  maka hasil *clustering* yang didapat juga lebih baik. Secara esensial, DBI menginginkan nilai

sekecil (non-negatif  $\geq 0$ ) mungkin untuk menilai baiknya *cluster* yang didapat. Indeks tersebut didapat dari rata-rata semua indeks *cluster*, dan nilai yang didapat bisa digunakan sebagai pendukung keputusan untuk menilai jumlah *cluster* yang cocok digunakan. DBI juga banyak digunakan untuk membantu *K-Means* dalam menentukan jumlah *cluster* yang tepat untuk digunakan karena biasanya *K-Means* belum bisa mengetahui berapa *cluster* yang digunakan untuk clustering data.

**METODE PENELITIAN**

Berikut adalah rancangan penelitian :



**Gambar 1.** Rancangan Penelitian

**a. Business Understanding**

Pada fase ini ada tiga tahap yang dilakukan, yaitu :

**a.1 Pemahaman terhadap tujuan bisnis**

Dalam aplikasinya pada perusahaan retail, perlu didahului dengan pemahaman terhadap tujuan bisnis yang akan menentukan pola yang akan dicari. Beberapa tujuan bisnis pada perusahaan retail yang berkaitan dengan identifikasi pelanggan berpotensi *churn* adalah :

1. Mengetahui siapa saja pelanggan yang loyal dan pelanggan terbaik
2. Mengetahui pelanggan mana yang berada dalam ambang *churn*
3. Meningkatkan dan mempertahankan jumlah konsumen terutama konsumen yang berpotensi untuk menjadi konsumen tetap (pelanggan)
4. Meningkatkan nilai penjualan terhadap melalui kegiatan *cross-selling* dan *up-selling*

**a.2 Penilaian Situasi**

UD. Mawar Sari merupakan perusahaan retail yang berdiri sejak Tahun 1976 dan berlokasi di Jln. Ahmad Yani No. 5, Subagan, Amlapura. UD. Mawar Sari menjual kebutuhan masyarakat sehari – hari secara eceran dan grosir. Sampai Tahun 2016, UD. Mawar Sari memiliki 2700 pelanggan, 4600 item barang, dan rata – rata jumlah transaksi penjualan sebanyak 1200 transaksi serta rata – rata beromzet 450 juta rupiah dalam sebulan. Sejak Tahun 2009, UD. Mawar Sari telah menggunakan sistem informasi berbasis komputer untuk menangani *point of sales* (POS) dan *inventory*.

**a.3 Menerjemahkan Tujuan Bisnis**

Pada tahapan ini dibutuhkan pemahaman terhadap tujuan bisnis dan menerjemahkan ke dalam pengolahan data. Salah satu tujuan pengolahan data dalam CRM untuk mendukung proses *customer identification* adalah identifikasi karakter pelanggan. Hasil identifikasi dapat digunakan oleh pihak manajemen untuk upaya

mengidentifikasi pelanggan, mempertahankan pelanggan dan meningkatkan nilai penjualan.

**b. Tinjauan Pustaka**

Pada tahapan ini peneliti melakukan tinjauan pustaka yaitu mengumpulkan dan mempelajari literature, buku, artikel, jurnal dan sebagainya yang diperoleh dari berbagai sumber seperti perpustakaan, internet, dll.

**c. Data Understanding**

Tahap ke tiga yaitu *Data Understanding*. Pada tahap ini dilakukan proses pengumpulan data, kemudian melakukan analisa data serta melakukan evaluasi kualitas data yang digunakan dalam penelitian ini. Dari proses perijinan data yang diajukan, diperoleh data transaksi pelanggan pada tahun 2015 sebanyak 30.765 *record* data, tahun 2016 sebanyak 64.008 *record* data dan pada tahun 2017 sebanyak 26.316 *record* data.

**d. Data Preparation**

Pada tahap ini data akan dipersiapkan sehingga mempermudah proses *mining*. Proses *preparation* ini mencakup yaitu :

- a. *Data Selection* : memilih data yang akan digunakan dalam proses data mining. Dalam proses ini dilakukan juga pemilihan atribut – atribut yang disesuaikan dengan proses data mining.
- b. *Data Preprocessing* : Memastikan kualitas data yang telah dipilih pada tahap data selection, pada tahap ini masalah yang harus dihadapi adalah noisy data dan missing values. Proses pembersihan data (*cleansing*) dilakukan dengan melakukan metoda – metoda query sederhana untuk menemukan anomaly – anomaly data yang bisa saja masih terdapat pada sistem.

**e. Membentuk fitur *recency, frequency, monetary***

Pada tahap ini dikelompokkan atribut – atribut atau field yang telah terpilih menjadi 1 tabel dengan cara melakukan denormalisasi dari basis data OLTP. Hasil pemilihan atribut – atribut yaitu :

**Tabel 1.** Pemilihan Atribut

Field	Keterangan
Kode Pelanggan	Kode Pelanggan
TglBeliAkhir	Menandakan <i>recency</i> , merupakan tanggal transaksi yang dilakukan oleh pelanggan
FrekuensiBeli	Menandakan <i>frequency</i> , merupakan jumlah transaksi selama periode yang ditentukan
TotalBeli	Menandakan <i>monetary</i> , merupakan jumlah uang selama periode yang ditentukan

Dalam proses *data preparation* dibangun suatu *data warehouse* untuk mempermudah proses *mining*. Data yang akan dilibatkan dalam penelitian ini adalah data yang berkaitan dengan transaksi penjualan. Data set yang digunakan dapat disesuaikan selama proses *data mining*, karena dimungkinkan untuk kembali ke fase sebelumnya jika ada informasi dan pengetahuan yang belum terungkap.

#### f. Pembentukan model RFM

Dalam pembentukan model RFM, diterapkan RFM skor satu sampai tiga (1-3) pada domain nilai *recency*, *frequency*, dan *monetary* di setiap pelanggan. Skor 3 adalah nilai tertinggi dan skor 1 adalah nilai terendah. Skor akhir dihitung dari kombinasi setiap skor pada atribut. Pemberian skor sendiri akan memudahkan proses segmentasi pelanggan.

Nilai *recency*, *frequency*, *monetary* dibagi menjadi 3 bagian dengan nilai 3, 2 dan 1. Nilai *recency* dihitung berdasarkan tanggal transaksi terakhir atau interval waktu transaksi terakhir dengan saat ini. Nasabah dengan tanggal transaksi terbaru mempunyai nilai 3 sedangkan nasabah dengan tanggal transaksi terjauh di masa lalu mempunyai nilai 1. Begitu juga dengan nilai *frequency*, nasabah yang sering bertransaksi mempunyai nilai *frequency* yang tinggi, yaitu 3. Sedangkan nasabah yang jarang bertransaksi mempunyai nilai 1. Nasabah yang mempunyai total nilai transaksi terbesar mempunyai nilai *monetary* yang tinggi, dengan nilai 3. Sebaliknya nasabah yang mempunyai total

nilai transaksi terkecil mempunyai nilai *monetary* yang rendah yaitu 1. Rentang skor dan domain nilai didapatkan dari hasil wawancara dengan manajemen UD. Mawar Sari dan hasil bagi 3 dari nilai atribut terkecil ke nilai atribut tertinggi. Skor dan domain nilai untuk atribut *recency*, *frequency*, *monetary* ditunjukkan dalam table berikut ini :

Setelah ditentukan skor dan domain nilai pada *recency*, *frequency*, dan *monetary* maka akan didapatkan skor akhir RFM dan penentuan label pelanggan.

Terdapat 27 (3×3×3) kombinasi skor akhir RFM. Skor akhir RFM yang tertinggi adalah 333 dan yang terendah adalah 111. Nasabah dengan skor akhir 333 merupakan nasabah yang dengan tingkat kelayakan yang tinggi sedangkan nasabah dengan skor akhir 111 merupakan nasabah dengan tingkat kelayakan yang rendah.

**Tabel 2.** Klasifikasi Atribut dengan Skor dan Domain Nilai

Atribut	Skor	Domain nilai
<i>Recency</i>	1	$r > 150$ hari
	2	$61 \text{ hari} \leq r \leq 150$ hari
	3	$0 \leq r \leq 60$ hari
<i>Frequency</i>	1	$0 < f \leq 30$
	2	$31 \leq f \leq 60$
	3	$f > 60$
<i>Monetary</i>	1	$0 < m \leq \text{Rp. } 1.500.000$
	2	$\text{Rp. } 1.500.000 < m \leq \text{Rp. } 3.500.000$
	3	$m > \text{Rp. } 3.500.000$

**Tabel 3.** Skor Akhir RFM dan Label Pelanggan

Kelas	Skor			Skor Akhir RFM	Label Pelanggan
	Recency	Frequency	Monetary		
K1	3	1	1	311	<i>Dormant D</i>
K2	3	1	2	312	<i>Dormant A</i>
K3	3	1	3	313	<i>Occational A</i>
K4	3	2	1	321	<i>Everyday D</i>
K5	3	2	2	322	<i>Golden D</i>
K6	3	2	3	323	<i>Superstar D</i>
K7	3	3	1	331	<i>Everyday A</i>
K8	3	3	2	332	<i>Golden A</i>
K9	3	3	3	333	<i>Superstar A</i>
K10	2	1	1	211	<i>Dormant E</i>
K11	2	1	2	212	<i>Dormant B</i>
K12	2	1	3	213	<i>Occational B</i>
K13	2	2	1	221	<i>Everyday E</i>
K14	2	2	2	222	<i>Golden E</i>
K15	2	2	3	223	<i>Superstar E</i>
K16	2	3	1	231	<i>Everyday B</i>
K17	2	3	2	232	<i>Golden B</i>
K18	2	3	3	233	<i>Superstar B</i>
K19	1	1	1	111	<i>Dormant F</i>
K20	1	1	2	112	<i>Dormant C</i>
K21	1	1	3	113	<i>Occational C</i>
K22	1	2	1	121	<i>Everyday F</i>
K23	1	2	2	122	<i>Golden F</i>
K24	1	2	3	123	<i>Superstar F</i>
K25	1	3	1	131	<i>Everyday C</i>
K26	1	3	2	132	<i>Golden C</i>
K27	1	3	3	133	<i>Superstar C</i>

**g. Klusterisasi dengan K-Means**

Proses klusterisasi segmentasi pelanggan menggunakan teknik *clustering* pada *data mining* dengan metode *K-Means Clustering*. Proses ini menggunakan data pada Juli 2015 – Juni 2016 dan data pada Mei 2016 – Mei 2017 hasil pemodelan RFM pada langkah sebelumnya.

Proses *clustering* data sebanyak 1719 *record* ini menggunakan perangkat lunak *Weka 3.6.10* dengan metode *K-Means Clustering*. *Clustering* diuji coba dengan berbagai jumlah *cluster* yaitu 2 sampai 5 *cluster* dengan bantuan tools *Weka 3.6.9 for Windows 32 bit* dilakukan dengan menetapkan nilai awal yaitu :

1. Jumlah *cluster*
2. Maksimum iterasi = 500

Selanjutnya untuk mendapatkan jumlah *cluster* terbaik, peneliti menggunakan metode uji validasi *cluster Davies-Bouldin Index*.

**h. Identifikasi Karakter Pelanggan Berbasis RFM**

Segmentasi adalah proses membagi pelanggan menjadi beberapa *cluster* dengan kategori loyalitas pelanggan untuk membangun strategi pemasaran. Segmentasi pelanggan dibagi menjadi 6 karakteristik berdasarkan nilai RFM

sebagai berikut :

1. *Superstar*

Pelanggan dengan *loyalty* yang tinggi dengan nilai *monetary*, frekuensi dan mempunyai transaksi paling tinggi.

2. *Golden*

Pelanggan yang mempunyai nilai *monetary* tertinggi kedua, frekuensi yang tinggi dan mempunyai rata-rata transaksi.

3. *Typical*

Pelanggan yang mempunyai rata-rata nilai *monetary* dan rata-rata transaksi

4. *Occational*

Pelanggan yang mempunyai nilai *monetary* terendah kedua setelah *dormant*, nilai *recency* paling rendah dan transaksi paling tinggi.

5. *Everyday*

Pelanggan yang memiliki peningkatan transaksi, transaksi yang rendah, dan mempunyai nilai *monetary* sedang sampai

rendah.

6. *Dormant*

Pelanggan yang mempunyai frekuensi dan *monetary* yang paling rendah dan *recency* yang paling rendah.

Pada penelitian ini menggunakan metoda RFM (*recency, frequency, monetary*) untuk mengidentifikasi karakter pelanggan dan menghasilkan kelas pelanggan. Fitur *recency, frequency, dan monetary* di dapatkan dari data pelanggan dan data transaksi pelanggan. *Recency* adalah kapan transaksi terakhir dilakukan. *Frequency* adalah tingkat keseringan pelanggan melakukan transaksi, dan *monetary* adalah besarnya nilai transaksi yang dilakukan.

**i. Evaluasi**

Dalam penelitian ini, uji validasi *cluster* dengan *Davies-Bouldin Index (DBI)* menggunakan perangkat lunak pengolah angka *Microsoft Excel 2010*. Nilai *Davies-Bouldin Index* yang kecil merupakan jumlah *cluster* yang baik. Semakin kecil nilai *Davies-Bouldin Index*, maka semakin optimal hasil *cluster*.

**HASIL PEMBAHASAN**

Proses segmentasi pelanggan dengan model RFM adalah persiapan data yang dibutuhkan untuk segmentasi pelanggan menggunakan pemodelan RFM (*recency, frequency, recency*) dan akan di klusterisasi menggunakan Metode *K-Means Clustering*. Proses *query database* UD. Mawar Sari menggunakan perangkat lunak *Microsoft Query* untuk ditransformasikan ke dalam *database* yang akan diolah.

Untuk mendapatkan atribut *recency, frequency, dan monetary*, hasil *query* diolah kembali menggunakan perangkat lunak *Microsoft Excel 2010* dengan fasilitas *Pivot Table*. Gambar 2 berikut menunjukkan pengolahan hasil *query* dalam bentuk *pivot table*.

Date of FISCAL TRANS	Sum of MEAS. TRANS
2015/07/01	145,400
2015/08/01	1,204,000
2015/09/01	1,232,800
2015/10/01	344,600
2015/11/01	431,600
2015/12/01	225,500
2016/01/01	31,500
2016/02/01	80,000
2016/03/01	1,422,800
2016/04/01	363,500
2016/05/01	966,000
2016/06/01	42,000
2016/07/01	3,738,500
2016/08/01	35,500
2016/09/01	222,100
2016/10/01	1,491,500
2016/11/01	1,622,800
2016/12/01	2,524,800
2017/01/01	1,796,700
2017/02/01	300,200
2017/03/01	4,783,700
2017/04/01	108,600
2017/05/01	2,587,700
2017/06/01	2,997,200
2017/07/01	675,200

**Gambar 2.** Pengolahan data dengan *pivot table*

Data yang disajikan oleh *Pivot Table* diolah kembali untuk mendapatkan skor RFM sesuai dengan domain nilai yang tercantum pada tabel 2. Gambar 3 di bawah ini menunjukkan data yang nilai setiap atributnya berupa skor RFM.

**Gambar 3.** Data hasil pemodelan RFM

**Tabel 4.** Hasil uji validasi cluster dengan DBI

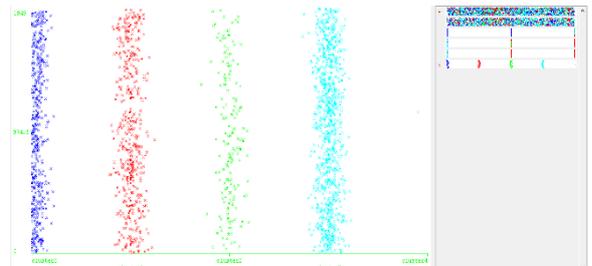
Data	Jumlah Cluster	Davies-Bouldin Index (DBI)
Data(Juli 2015 – Juni 2016)	2	0.918475
	3	0.71335231
	4	0.415896659
	5	0.400261
Data (Mei 2016-Mei 2017)	2	0.719383087
	3	0.492880934
	4	0.929543034
	5	0.929981698

Untuk data pada bulan Juli 2015 – Juni 2016 hasil klusterisasi terbaik adalah 5 kluster, sedangkan data pada bulan Mei 2016 – Mei 2017 hasil klusterisasi terbaik adalah 3 kluster. Berikut table hasil klusterisasi terbaik.

**Tabel 5.** Hasil 5 (lima) kluster

No Cluster	Jumlah anggota (%)	Jumlah anggota	R	F	M
1	4%	61	3	1	1,9836
2	12%	200	1	1.005	1.005
3	19%	324	2	1.1344	1.0309
4	43%	744	3	1.1344	1
5	23%	390	3	2.5641	2.3744

Hasil proses *clustering* dengan 5 cluster ditunjukkan berupa grafik. Proses *clustering* dihentikan pada iterasi ke- 5 dengan nilai *sum of squared errors* = 2055.5147692704295. Waktu yang dibutuhkan untuk membangun model adalah 0.06 detik. Grafik hasil *clustering* dengan 5 cluster ditampilkan pada Gambar 4.

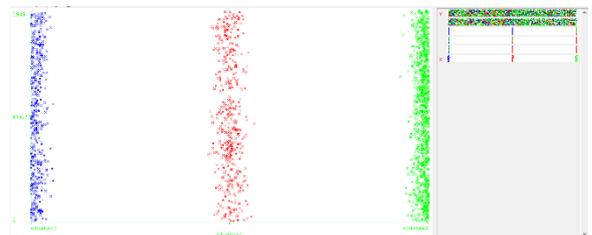


**Gambar 4.** Grafik Hasil *clustering* Data Uji dengan 5 cluster

**Tabel 6.** Hasil 3 (tiga) kluster

No Cluster	Jumlah anggota (%)	Jumlah anggota	R	F	M
1	22%	432	2	1.002	1.011
2	23%	453	3	2.476	1.408
3	55%	1065	3	1.096	1.063

Hasil proses *clustering* dengan 3 cluster ditunjukkan berupa grafik. Proses *clustering* dihentikan pada iterasi ke- 3 dengan nilai *sum of squared errors* = 2076.206949302488. Waktu yang dibutuhkan untuk membangun model adalah 0.03 detik. Grafik hasil *clustering* dengan 3 cluster untuk kelompok *recency* ditampilkan pada gambar 5.



**Gambar 5.** Grafik Hasil *clustering* Data Uji dengan 3 cluster

**SIMPULAN**

Dari proses segmentasi pelanggan menggunakan pemodelan RFM (*Recency, Frequency, Monetary*) dapat diketahui karakteristik pelanggan UD. Mawar Sari. Dari proses segmentasi tersebut dapat diketahui kelas pelanggan yang dimiliki oleh UD. Mawar Sari yaitu kelas *Dormant, Everyday, Golden* dan *Superstar*.

Uji coba pengelompokkan kelas pelanggan dengan *K-Means Clustering* pada beberapa jumlah cluster yaitu 2 *cluster* sampai dengan 5 *cluster*, setelah diuji dengan *Davies-Boundin Index* mendapatkan jumlah 5 *cluster* sebagai pengelompokkan kelas pelanggan yang terbaik. *Cluster* ke-1 ditempati kelas pelanggan *Dormant D* dan *Dormant A*, *Cluster* ke-2 ditempati kelas pelanggan berpotensi *churn* yaitu *Dormant F*, *Cluster* ke-3 ditempati kelas pelanggan *Dormant B*, *Dormant E* dan *Everyday E*, *Cluster* ke-4 ditempati kelas pelanggan *Dormant D* dan *Everyday D*, *Cluster* ke-5 ditempati kelas pelanggan potensial dengan tingkat loyalitas tinggi yaitu *Golden D*, *Golden A*, *Superstar D*, *Superstar A* dan *Everyday A*.

Pelanggan UD. Mawar Sari yang berpotensi *churn* adalah kelas pelanggan yang berada di *cluster* ke-2 dan *cluster* ke-3 yaitu kelas pelanggan *Dormant B*, *Dormant E*, *Dormant F* dan *Everyday E*.

**DAFTAR PUSTAKA**

- [1] a a G. B. Ariana, "Customer Segmentation Dengan Metode Self Organizing Map ( Studi Kasus : UD . Fenny)," *Neuron*, vol. 2, no. 1, pp. 1–7, 2011.
- [2] W. Buckinx and D. Van Den Poel, "Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting," *Eur. J. Oper. Res.*, vol. 164, no. 1, pp. 252–268, 2005.
- [3] J. Burez and D. Van den Poel, "CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services," *Expert Syst. Appl.*, vol. 32, no. 2, pp. 277–288, 2007.
- [4] A. Chorianopoulos, *Effective CRM using Predictive Analytics*. Wiley, 2009.
- [5] Y. Liu and Y. Zhuang, "Research Model of Churn Prediction Based on Customer Segmentation and Misclassification Cost in the Context of Big Data," *J. Comput. Commun.*, vol. 3, no. 3, pp. 87–93, 2015.
- [6] V. L. M. Oliviera, "Analytical Customer Relationship Management in Retailing Supported by Data Mining Techniques," University of Porto, 2012.
- [7] V. L. M. Oliviera, "Predicting Partial Customer Churn using Markov for Discrimination for Modeling First Purchase Sequence," University of Porto, 2012.
- [8] Wardani, N. W. (2020). *Penerapan Data Mining Dalam Analytic CRM*. Yayasan Kita Menulis.
- [9] Wardani, N. W., & Ariasih, N. K. (2019). Analisa Komparasi Algoritma Decision Tree C4. 5 dan Naïve Bayes untuk Prediksi Churn Berdasarkan Kelas Pelanggan Retail. *International Journal of Natural Science and Engineering*, 3(3), 103-112.
- [10] Wardani, N. W., Dantes, G. R., & Indrawan, G. (2018). Prediksi Customer Churn dengan Algoritma Decision Tree C4. 5 Berdasarkan Segmentasi Pelanggan untuk Mempertahankan Pelanggan pada Perusahaan Retail. *Jurnal RESISTOR (Rekayasa Sistem Komputer)*, 1(1), 16-24.
- [11] Darmansah, D., & Wardani, N. W. (2020). Analisa Penyebab Kerusakan Tanaman Cabai Menggunakan Metode K-Means. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 7(2), 126-134.