

Segmentasi Pelanggan Menggunakan Metode *K-Means Clustering* Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang)

Aulia Dewi Savitri¹, Fitra Abdurrachman Bachtiar², Nanang Yudi Setiawan³

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹auliadewis08@gmail.com, ²fitra.bachtiar@ub.ac.id, ³nanang@ub.ac.id

Abstrak

Belle Crown merupakan salah satu klinik kecantikan baru di Malang yang belum menerapkan strategi CRM (*Customer Relationship Management*) dengan memberikan pelayanan yang berbeda pada setiap pelanggan. Segmentasi merupakan proses untuk mengetahui karakteristik pelanggan dengan kesamaan tertentu, sehingga mempermudah penggalan informasi terkait pelanggan yang *profitable*. Perilaku bisnis pelanggan dapat dilihat dari *Recency* (rentang waktu transaksi terakhir), *Frequency* (jumlah transaksi), dan *Monetary* (rupiah yang dikeluarkan) atau dikenal dengan analisis RFM (*Recency, Frequency, Monetary*). Salah satu metode pengelompokan yaitu *K-Means* yang dapat digunakan untuk melakukan segmentasi tersebut. Hasil grafik dari metode *Elbow* digunakan untuk membantu dalam menentukan jumlah segmen secara intuitif dalam penerapan metode *K-Means*. Data yang digunakan dalam penelitian ini adalah riwayat transaksi pelanggan pada periode Mei hingga Oktober 2017 sejumlah 21.513 transaksi dan 4716 pelanggan. Pada penelitian ini menghasilkan dua segmen dan tiga segmen. Analisis berdasarkan nilai RFM menunjukkan bahwa peringkat pertama merupakan kelompok pelanggan yang *profitable* karena memiliki nilai RFM yang lebih besar dibanding segmen lainnya. Luaran dari penelitian ini adalah visualisasi *dashboard* dengan beberapa grafik yang memuat hasil segmentasi pelanggan berdasarkan nilai RFM pada layanan-layanan yang diberikan oleh Belle Crown.

Kata kunci: segmentasi pelanggan, *clustering*, analisis RFM, *K-Means*

Abstract

Belle Crown is one of new aesthetic centers in Malang has not applied CRM strategy (Customer Relationship Management) by giving different service for all of its costumers. Segmentation is a process undergo to identify costumers with similar characteristics, therefore, it can help to explore more information on profitable costumers. The costumer's business behaviour could be seen from Recency (last transaction range), Frequency (the number of transactions), and Monetary (the amount of money spent) or it is known as RFM (Recency, Frequency, Monetary). One of data clustering method is K-Means that is used to do the segmentation. The graphics result from Elbow method is used to determine the number of segments intuitively during the application of K-Means method. The data used in this research is transaction history taken from May-October 2017 and it includes 21.513 transactions and 4716 costumers. In its application, the research results two kinds of segments including 2 segments and 3 segments. The analysis based on RFM value showed that the first rate is the profitable customer as it has bigger RFM compared to other segments. The superficial of this research is to produce dashboard visualization as the result of costumers segmentation with some graphics based on RFM value of Belle Crown's service.

Keywords: costumers segmentation, clustering, RFM Analysis, *K-Means*

1. PENDAHULUAN

Belle Crown adalah salah satu klinik kecantikan yang tergolong baru di Kota Malang. Adapun kegiatan utamanya adalah memberikan

pelayanan yang optimal guna memenuhi kebutuhan pelanggan. Seperti yang diketahui di era modern ini, seiring perkembangan teknologi, ekonomi, dan budaya, *treatment* kecantikan menjadi salah satu kebutuhan penting bagi

sebagian masyarakat khususnya kaum hawa. Hal tersebut dibuktikan bahwa dalam 10 tahun terakhir industri kecantikan dan perawatan pribadi di Indonesia bertumbuh rata-rata 12% dengan nilai pasar mencapai sebesar 33 triliun rupiah di tahun 2016. Bahkan di tahun 2020, industri kecantikan di Indonesia diprediksi akan mengalami pertumbuhan paling besar dibandingkan dengan negara-negara lain di Asia Tenggara (Lestari, 2017). Kondisi ini menyebabkan para pelaku bisnis bersaing untuk memberikan pelayanan terbaik sesuai kebutuhan konsumen. Salah satu cara yakni melalui penerapan strategi *Customer Relationship Management* (CRM). Salah satu konsep CRM adalah pelayanan yang berbeda pada setiap pelanggan sesuai dengan kebutuhan dan informasi karakteristik yang tercatat pada perusahaan (Rainer and Cagielski, 2011). Hal tersebut bertujuan meningkatkan pelayanan pada pelanggan yang disebut dengan strategi segmentasi pelanggan.

Akan tetapi, muncul beberapa permasalahan pada Belle Crown Malang diantaranya adanya sekumpulan data transaksi dengan informasi berupa waktu transaksi dan besaran rupiah pembayaran transaksi pelanggan belum diolah secara baik untuk mendapatkan informasi yang dapat meningkatkan nilai bagi perusahaan. Selain itu, belum adanya pemetaan pelanggan yang menyebabkan penerapan perlakuan yang sama pada seluruh pelanggan misalnya dalam hal penentuan promo, pemberian hadiah, maupun pelayanan. *Software* yang digunakan juga hanya sebatas untuk menyimpan data, sehingga belum terdapat visualisasi yang menggambarkan keadaan bisnis guna mempermudah perusahaan dalam mengambil kebijakan strategis. Hal tersebut, menyebabkan adanya sebagian pelanggan tidak dapat dipertahankan retensinya dan kesulitan perusahaan dalam menganalisis pelanggan yang loyal. Sedangkan seiring meningkatnya persaingan bisnis yang semakin ketat, banyak perusahaan hampir di semua bidang yang kemudian tidak lagi memfokuskan aktivitas pemasaran mereka hanya pada pencarian pembeli baru, namun sudah berubah menjadi pada usaha untuk mempertahankan dan meningkatkan kesetiaan pelanggan lama (Gaffar, 2007).

Dalam penelitian ini perlu untuk melakukan segmentasi pelanggan berdasarkan profil dari masing-masing pelanggan kemudian menganalisis hasil segmentasinya untuk

diketahui pelanggan yang *profitable* maupun sebaliknya. Untuk melakukan *profiling* pada pelanggan maka diperlukan sebuah model yang memberikan gambaran segala aktivitas pelanggan, kebutuhan, keinginan, dan juga konsentrasi terhadap produk dan layanan perusahaan. Model yang umum dalam mengelompokkan pelanggan adalah model *Recency, Frequency, Monetary* (RFM), yaitu melakukan mengelompokkan pelanggan berdasarkan interval waktu kunjungan terakhir pelanggan, frekuensi kunjungan, dan besaran nilai yang dikeluarkan sebagai royalti perusahaan (Aggelis et al., 2005 dan Chen et al., 2009). Metode yang sering digunakan dalam pengelompokan pelanggan adalah *data mining* terutama dengan teknik *clustering*.

Metode *clustering* digunakan untuk melakukan segmentasi pelanggan pada Belle Crown adalah *K-Means clustering*. Alasan metode ini digunakan adalah metode ini merupakan metode interaktif yang mudah diinterpretasikan, diterapkan, dan bersifat dinamis pada data yang tersebar (Hughes, 1994).

Pada beberapa penelitian terdahulu oleh Atyanto et al. (2011) dan Angelie (2017) telah dilakukan penelitian terkait segmentasi dengan melibatkan tiga buah variabel, yaitu R, F, M. Dari penelitian tersebut dijelaskan bahwa terdapat kekurangan bahwa pengelompokan dengan hanya menggunakan tiga variabel dianggap belum merepresentasikan karakteristik pelanggan yang sebenarnya. Oleh karena itu, pada penelitian ini melibatkan sejumlah atribut tambahan selain R, F dan M yakni dengan membandingkan hasil R, F, dan M dengan layanan yang ditawarkan pada Klinik Kecantikan Belle Crown Malang. Kemudian hasilnya divisualisasikan ke dalam bentuk grafik yang mudah untuk dipahami.

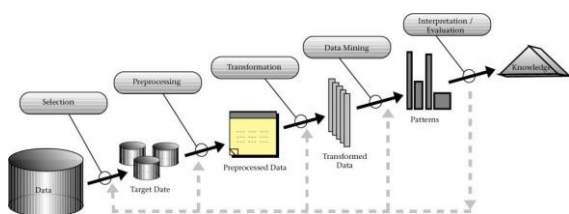
Relationship Marketing mengalami perkembangan dalam hal konsep menjadi *Customer Relationship Management* (CRM). CRM merupakan kombinasi dari proses dan teknologi yang bertujuan memahami pelanggan dalam hal perbedaan produk dan jasa yang digunakan. CRM memungkinkan pemberian layanan yang unggul pada pelanggan berdasarkan penggunaan informasi yang efektif dari pelanggan (Kotler and Keller, 2007). Menurut Kotler dan Keller (2007:35), terdapat empat aktivitas CRM yaitu mengidentifikasi (*identify*), mengakuisisi (*acquire*), mempertahankan (*retain*), mengembangkan (*develop*).

Sedangkan analisis RFM merupakan proses analisis perilaku pelanggan. Proses ini biasanya digunakan dalam pemasaran *database* dan pemasaran langsung. Untuk menentukan segmentasi pelanggan maka digunakan model RFM berdasarkan tiga variabel yaitu *Recency*, *Frequency*, *Monetary* (Tsipitsis dan Chorianopoulos, 2009).

1. *Recency*, yaitu rentang terakhir kali transaksi dilakukan. Semakin kecil rentangnya, maka nilai *R* semakin besar.
2. *Frequency*, yaitu jumlah transaksi dalam satu periode. Semakin banyak *frequency*, maka nilai *F* semakin besar.
3. *Monetary*, yaitu nilai pelanggan berupa uang yang dikeluarkan selama transaksi. Semakin banyak jumlah besaran uang yang dikeluarkan pelanggan pada periode tersebut maka nilai *M* semakin besar.

Menurut Cheng dan Chen (2009), semakin besar nilai *R* dan *F* maka kemungkinan pelanggan akan melakukan transaksi kembali dengan perusahaan tersebut. Selain itu semakin besar nilai *M*, maka kecenderungan pelanggan dalam memberikan respon kepada produk dan layanan perusahaan.

Sekumpulan data dengan jumlah yang besar dapat digali informasinya dengan istilah *data mining* dan *knowledge discovery in databases* (KDD). Kedua istilah tersebut memiliki arti yang berbeda namun saling terkait. Han dan Kamber (2006) menyatakan bahwa *data mining* adalah salah satu tahapan dalam proses KDD. Berikut adalah tahapan KDD pada Gambar 1.



Gambar 1. Tahapan KDD
(Han dan Kamber, 2006)

Dalam proses transformasi dibutuhkan normalisasi dilakukan dengan menyesuaikan parameter RFM yang telah dibuat (Khajvand dan Tarokh, 2011). Metode ini merupakan metode normalisasi dengan cara melakukan transformasi atribut numerik dalam skala yang lebih kecil yaitu dengan batas terendah 0 dan batas tertinggi 1 (Junaedi, 2011). Berikut adalah perhitungan normalisasi *Min-Max* seperti pada persamaan (1).

$$x' = \frac{x - \min_a}{\max_a - \min_a} (\text{newmax} - \text{newmin}) + \text{newmin} \dots (1)$$

Dimana:

x' = nilai setelah dinormalisasi

x = nilai asli yang akan dinormalisasi

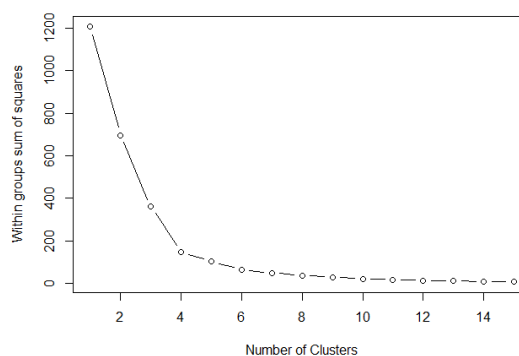
\min_a = nilai minimal setiap variabel.

\max_a = nilai maksimal setiap variabel.

newmax = rentang maksimal x dengan nilai 1.

newmin = rentang minimal x dengan nilai 0.

Untuk menentukan jumlah *cluster* menggunakan metode *Elbow*. Metode ini yang digunakan dalam analisis *cluster* untuk interpretasi dan uji performa tingkat konsistensi jumlah *cluster* yang tepat dengan melihat nilai SSE (Bholowalia dan Kumar, 2014). Pada titik tertentu akan terjadi grafik penurunan secara drastis dengan sebuah lekukan yang disebut dengan kriteria siku. Nilai itu kemudian menjadi nilai k atau jumlah *cluster* yang terbaik (Bholowalia dan Kumar, 2014). Berikut adalah grafik metode *Elbow* pada Gambar 2.



Gambar 2. Grafik Metode *Elbow*
(Bholowalia dan Kumar, 2014)

Adapun menurut Bholowalia dan Kumar (2014) algoritma metode *Elbow* dalam menentukan nilai k pada *K-Means* adalah :

1. Inisialisasi $k = 1$
2. Mulai
3. Kenaikan nilai k
4. Mengukur SSE
5. Jika di suatu titik SSE turun drastis
6. Hal tersebut menunjukkan nilai k yang benar
7. Selesai.

Data clustering menggunakan metode *K-Means* ini secara umum dilakukan dengan algoritma dasar sebagai berikut (Tan dan Kumar, 2005) :

1. Tentukan jumlah kluster k .
 - a. Pemberian nilai secara random.
 - b. Pengambilan sampel awal data.

- c. Penentuan nilai awal hasil kluster hirarki dengan jumlah kluster yang sesuai dengan penentuan awal.
2. Inisialisasi k pusat kluster.
3. Tempatkan objek ke kluster terdekat dengan *Euclidean Distance*.
 - a. Berdasarkan kedekatannya dengan pusat kluster.
 - b. Dihitung jarak tiap data ke tiap pusat kluster.
 - c. Jarak paling dekat tersebut akan menjadi penentu masuk kluster mana.
4. Perhitungan kembali pusat kluster.
 - a. Menghitung nilai rata rata data dalam kluster tertentu.
 - b. Pada kasus tertentu digunakan median, karena median tidak sensitif terhadap *outlier* data.
 - c. Ulangi langkah 3 dengan memakai pusat kluster baru. Jika pusat kluster tidak berubah lagi maka proses pengklasteran dihentikan.

Pada proses *K-Means*, nilai *within sum of squares* menunjukkan variansi *cluster*, semakin kecil nilai *within sum of squares* maka *cluster* yang terbentuk semakin baik (Calinski dan Harabasz, 1974).

Dalam uji performa membutuhkan tiga metode dalam *package* CIVValid() yakni *Dunn Index*, *Silhouette Width*, dan *Connectivity*. *Dunn Index* merupakan metrik untuk memvalidasi hasil pengelompokan dengan mengukur diameter *cluster* dan jarak antara dua *cluster*. Dengan mengukur *Dunn Index* dapat diketahui jumlah *cluster* yang optimum (Khajvand dan Tarokh, 2011). *Dunn Index* dihitung dengan mengukur rasio jarak terbesar antar *cluster* dengan jarak terkecil di dalam *cluster* (Bhatia, 2012). Semakin tinggi nilai *Dunn Index* maka semakin baik *cluster* yang terbentuk (Brock, 2011). Berikut adalah perhitungan *Dunn Index* menurut Saitta dkk. (2007) pada persamaan (2).

$$DU = \min_{i=1, \dots, k} \left\{ \min_{j=i+1, \dots, k} \left(\frac{\text{diss}(c_i, c_j)}{\max_{m=1, \dots, k} (\text{diam}(c_m))} \right) \right\} \dots (2)$$

Dimana:

DU = *Dunn Index*.

k = jumlah k *cluster*.

(c_i, c_j) = jarak *cluster* i dan *cluster* j .

$\text{diam}(c_m)$ = diameter *cluster* i .

Silhouette width merupakan hasil rata-rata dari nilai *Silhouette* setiap observasi. *Silhouette*

mengukur derajat kepercayaan penempatan dalam *clustering*, jika nilai *Silhouette* mendekati 1 maka penempatan *cluster* sudah sesuai, akan tetapi jika nilainya mendekati -1 maka penempatannya tidak sesuai (Brock, 2011). Berikut persamaan untuk mencari *Silhouette Width* menurut Brock dkk. (2011) pada persamaan (3).

$$S(i) = \frac{b_i - a_i}{\max(b_i, a_i)} \dots (3)$$

Dimana:

a_i = rata-rata jarak i dengan observasi lain dalam satu *cluster*.

b_i = rata-rata jarak i dengan observasi pada *cluster* terdekat.

Sedangkan metode *Connectivity* merepresentasikan kekuatan hubungan antar anggota *cluster*. Nilai *Connectivity* memiliki rentang antara nol dan ∞ . Semakin rendah nilai *Connectivity*, maka jumlah *cluster* semakin optimal (Brock, 2011). Berikut adalah persamaan *Connectivity* menurut Brock dkk. (2011) pada persamaan (4).

$$Conn = \sum_{i=1}^N \sum_{j=1}^L x_{i, nni(j)} \dots (4)$$

Dimana:

$nni(j)$ = tetangga terdekat dari j dari pengamatan di i .

$x_{i, nni(j)}$ = mendekati nol jika i dan j dalam satu *cluster* dan 1 jika sebaliknya.

L = parameter yang menentukan jumlah tetangga

Untuk memverifikasi hasil *cluster Euclidean distance* merupakan persamaan untuk menghitung jarak antar data dengan menghitung akar perbedaan data yang berpasangan. Jarak tersebut digunakan untuk menentukan *cluster* dari anggota pada proses *K-Means Clustering* (Singh dkk., 2013). Berikut adalah persamaan *Euclidean Distance* menurut Singh dkk. (2013) pada persamaan (5).

$$\text{dist}((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2} \dots (5)$$

Dimana :

$\text{Dist}()$ = *Euclidean Distance*

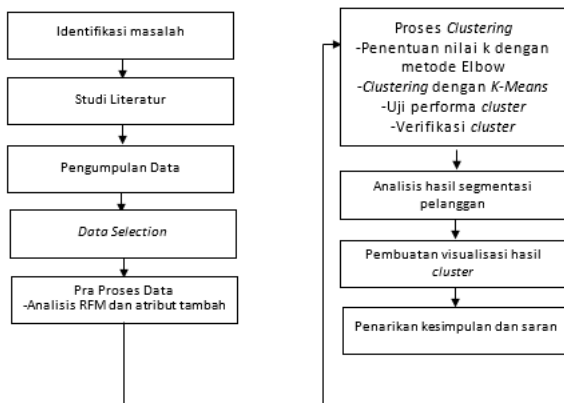
(x, y) = koordinat data

(a, b) = koordinat *centroid*

2. METODOLOGI

Adapun alur yang akan dilakukan dalam

penelitian ini dapat dilihat dalam Gambar 3:



Gambar 3. Alur Penelitian

2.1 Identifikasi Masalah

Melakukan analisis terhadap objek yang ditentukan yaitu Klinik Kecantikan *Belle Crown* berdasarkan kondisi objek penelitian dan masalah segmentasi pelanggan.

2.2 Studi Literatur

Mencari teori – teori yang relevan terhadap penelitian sejenis yang pernah dilakukan sebelumnya yaitu oleh Atyanto et al. (2011) dan Angelie (2017) telah dilakukan penelitian terkait segmentasi dengan melibatkan tiga buah variabel, yaitu RFM dan metode segmentasi menggunakan *K-Means Clustering*.

2.3 Pengumpulan Data

Pengumpulan data dilakukan di Belle Crown Malang. Jenis data yang diperlukan yaitu data profil pelanggan klinik kecantikan *Belle Crown* yang bersifat *anonymous* atau bukan data *credential* pasien seperti atribut ID dan nama (dikarenakan aturan kode etik), data transaksi pelanggan 6 bulan terakhir tahun 2017, data preferensi RFM, dan data pendukung lainnya yang diekstraksi dari *database* dalam format .xls. Jumlah data yang didapatkan adalah 21.513 transaksi dan 4716 pelanggan.

2.4 Data Selection

Proses *data selection* pada data untuk memilih variabel preferensi RFM. Variabel *Recency* diperoleh dengan perhitungan selisih terakhir kali transaksi pelanggan dengan tanggal penelitian yakni 07 November 2017. Variabel *Frequency* dihitung dengan menjumlahkan transaksi pelanggan dalam satu periode. Sedangkan variabel *Monetary* dihitung dengan menjumlahkan besaran rupiah yang dikeluarkan dalam transaksi. Kemudian data diubah ke dalam

format .csv untuk dilakukan *pre processing data*.

2.5 Pre Processing Data

Sebelum proses *clustering*, perlu dilakukan *pre processing data* melalui *data cleansing*, reduksi *oulier*, dan transformasi data dengan proses *Logarithmic* untuk mentransformasi rentang nilai tiap variabel menjadi lebih kecil dan normalisasi Min-Max untuk mengubah rentang nilai tiap variabel menjadi 0 sampai dengan 1. Proses *pre processing* hingga proses *clustering* dilakukan menggunakan *software R*.

2.6 Clustering

Proses utama yaitu *clustering* selanjutnya dilakukan melalui tiga tahapan yaitu menentukan nilai *k* dengan metode *Elbow* secara *intuitive*, proses *K-Means Clustering* dan yang terakhir adalah uji performa *cluster* dengan metode hitung manual *SSE*, *Dunn Index*, *Silhouette Width*, dan *Connectivity*. *Input* dari proses ini adalah data pada variabel RFM. Setelah semua tahapan metode telah diterapkan, maka dilakukan analisis hasil *cluster* dengan melakukan denormalisasi data yakni data yang telah dinormalisasi dan ditransformasi dikembalikan ke nilainya yang semula. Tahapan ini dilakukan untuk mempermudah analisis *cluster* dengan membandingkan pada variabel layanan.

2.7 Pembuatan Visualisasi Cluster

Pada tahap visualisasi *cluster*, hasil *script* pada *software R* dimasukkan ke dalam *R query* pada *software Power BI*. *Power BI* membantu memvisualisasikan hasil *cluster* dalam bentuk grafik tiap segmen yang diperoleh pada proses sebelumnya. Selain grafik, ditampilkan juga anggota per *cluster*-nya beserta informasi hasil analisis yang dilakukan pada tahap sebelumnya. *Input* proses ini *cluster* yang dihasilkan oleh tools *R* dan *output* berupa *dashboard* yang menampilkan grafik hasil *clustering* berupa segmentasi pelanggan yang telah diverifikasi oleh pemilik Belle Crown Malang melalui kuisioner *System Usability Scale (SUS)*.

3. HASIL DAN PEMBAHASAN

3.1 Analisis RFM

Mengekstraksi data pelanggan berdasarkan tiga variabel RFM setiap pelanggan dalam waktu 6 bulan terakhir di tahun 2017. Setiap pelanggan diakumulasi masing-masing nilai RFM dalam periode 6 bulan terakhir seperti

terlihat pada potongan data pada Tabel 1 sampaidengan Tabel 3.

.Tabel 1. Tabel *Recency* Pelanggan

id_customer	Recency (hari)
001.2011.000649	32
001.2011.000679	72
001.2011.000684	139
001.2011.000705	89
001.2011.000709	20

.Tabel 2. Tabel *Frequency* Pelanggan

id_customer	Frequency (kali)
001.2011.000649	11
001.2011.000679	3
001.2011.000684	1
001.2011.000705	6
001.2011.000709	8

Tabel 3. Tabel *Monetary* Pelanggan

id_customer	Monetary (rupiah)
001.2011.000649	1822000
001.2011.000679	310000
001.2011.000684	35000
001.2011.000705	905000
001.2011.000709	1145000

3.2 Pra Proses Data

Terdapat tiga proses yaitu *data transformation* dengan *Logarithmic*, *data cleansing*, dan normalisasi Min-Max. Data hasil pra proses data terlihat pada Tabel 4 sampaidengan Tabel 6.

Tabel 4. Tabel Hasil Pra Proses Data *Recency*

id_customer	Recency
001.2011.000649	0.561619
001.2011.000679	0.305183
001.2011.000684	0.097167
001.2011.000705	0.238152
001.2011.000709	0.710246

Tabel 5. Tabel Hasil Pra Proses Data *Frequency*

id_customer	Frequency
001.2011.000649	0.669145
001.2011.000679	0.306574
001.2011.000684	0
001.2011.000705	0.5
001.2011.000709	0.580279

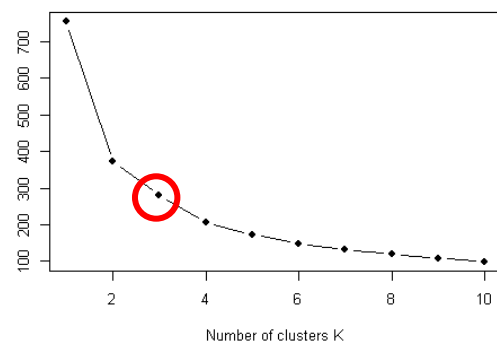
Tabel 6. Tabel Hasil Pra Proses Data *Monetary*

id_customer	Frequency
001.2011.000649	0.74846760
001.2011.000679	0.43937308

001.2011.000684	0.05870699
001.2011.000705	0.62634671
001.2011.000709	0.66739805

3.3 Penentuan Jumlah Cluster

Penentuan nilai k menggunakan *Elbow Method*. Pada plot tersebut didapatkan titik siku yang terbentuk diantara titik dua dan empat, setelah titik 3 sudah tidak lagi terjadi penurunan yang signifikan secara *intuitive*, sehingga dapat disimpulkan bahwa jumlah *cluster* menurut metode Elbow yaitu sebanyak 3 *cluster*. Akan tetapi jika dihitung kembali dengan perhitungan uji performa, didapatkan hasil bahwa jumlah *cluster* yang optimal adalah sejumlah 2 *cluster* sehingga dipilih titik kedua untuk dilakukan *clustering* ulang.



Gambar 4. Hasil Metode *Elbow*

3.4 K-Means Clustering

Hasil dari proses *clustering K-Means* berupa informasi yang menunjukkan jumlah anggota setiap *cluster*, titik pusat atau *centroid*, dan nilai performa *cluster* berdasarkan *within cluster sum of squares*.

a. Hasil K-Means 2 Cluster

Jumlah anggota setiap *cluster* tidak sama banyak. *Cluster* yang pertama sebanyak 2492 anggota dan *cluster* yang kedua sebesar 2215 anggota. Terdapat dua titik *centroid* yang saling berjauhan. Berikut adalah data hasil 2 *cluster* pada Tabel 7.

Tabel 7. Tabel Hasil 2 Cluster

Id_customer	Recency	Frequency	Monetary	CI
001.2011.000649	0.561619	0.669145	0.74846760	2
001.2011.000679	0.305183	0.306574	0.43937308	1
001.2011.000684	0.097167	0	0.05870699	1
001.2011.000705	0.238152	0.5	0.62634671	2
001.2011.000709	0.710246	0.580279	0.66739805	2

b. Hasil K-Means 3 Cluster

Jumlah anggota setiap *cluster* tidak sama banyak. *Cluster* yang pertama

sebanyak 1499 anggota, *cluster* yang kedua sebesar 1617 anggota, dan *cluster* ketiga 1591 anggota. Pada 3 segmen ini terdapat titik-titik *centroid* yang saling berjauhan terutama diantara titik *centroid cluster* 1 dan titik *centroid* di *cluster* 3. Berikut adalah data hasil 3 *cluster* pada Tabel 8.

Tabel 8. Tabel Hasil 3 *Cluster*

Id_customer	Recency	Frequency	Monetary	Cl
001.2011.000649	0.561619	0.669145	0.74846760	1
001.2011.000679	0.305183	0.306574	0.43937308	2
001.2011.000684	0.097167	0	0.05870699	3
001.2011.000705	0.238152	0.5	0.62634671	2
001.2011.000709	0.710246	0.580279	0.66739805	1

3.5 Uji Performa Menggunakan SSE

Pada pengujian ini mencari performa setiap jumlah *cluster* yang disesuaikan dengan rentang nilai pada metode *Elbow*. Berikut adalah hasil dari uji performa SSE pada Tabel 9.

Tabel 9. Hasil Uji SSE

Jumlah k	SSE	Selisih
1	755.47151	
2	371.39747	384.074
3	280.39807	90.9994
4	206.60417	73.7939
5	173.17318	33.43099
6	147.99011	25.18307
7	131.49566	16.49445
8	119.92141	11.57425
9	108.33301	11.5884
10	99.02402	9.30899

Berdasarkan Tabel 6.13 penurunan drastis terlihat pada titik 1 *cluster* dan 2 *cluster*. Sedangkan pada titik berikutnya mengalami penurunan yang stabil. Maka nilai k yang digunakan adalah 2. Sehingga perlu dilakukan *clustering* ulang dengan nilai k=2.

3.6 Uji Performa Menggunakan CValid()

Pada uji performa ini, hasil yang digunakan yakni nilai jumlah *cluster* optimal yang paling banyak muncul pada ketiga metode uji performa. Pada hasil *Connectivity* memilih jumlah *cluster* sebanyak 2 segmen karena memiliki nilai terendah diantara yang lain yaitu sebesar 162.5881, hasil *Dunn Index* memilih jumlah *cluster* sebanyak 5 segmen karena memiliki nilai tertinggi diantara yang lain yaitu sebesar 0.0085, sedangkan hasil *Silhouette* memilih jumlah *cluster* sebanyak 2 segmen karena memiliki nilai tertinggi diantara yang lain yaitu sebesar 0.4242. Dari data tersebut didapatkan hasil yang sama

pada uji *Connectivity* dan *Silhouette* yakni pada nilai k samadengan 2 yang berarti jumlah *cluster* sebanyak 2 merupakan *cluster* yang optimal. Berikut hasil data uji performa pada Tabel 10.

Tabel 10. Hasil Uji CValid()

Cluster	Validation Measure		
	Connectivity	Dunn Index	Silhouette Width
2	162.58	0.0013	0.4242
3	249.89	0.0051	0.3428
4	303.183	0.0059	0.3627
5	357.78	0.0085	0.3334
6	415.21	0.0067	0.3376
7	473.09	0.0079	0.3331
8	611.4	0.0040	0.3175
9	611.18	0.0037	0.3141
10	679.86	0.0053	0.3126

3.7 Analisis Clustering

1. Analisis 2 *Cluster*

Tabel 11. Hasil Rata-rata 2 *Cluster*

Ckuster	Recency	Frequency	Monetary	Rank
1	105	2	291660	2
2	34	6	1167493	1

Berdasarkan hasil rata-rata pada Tabel 11, pada 2 *cluster*, *cluster* 1 menempati peringkat kedua dengan rata-rata *recency* sebesar 105 hari atau transaksi terakhir yakni pada bulan Juli, rata-rata *frequency* sebanyak 2 kali, dan rata-rata *monetary* sebanyak 291660 rupiah, sedangkan *cluster* 2 menempati peringkat pertama dengan rata-rata *recency* sebesar 34 hari atau transaksi terakhir yakni pada bulan Oktober, rata-rata *frequency* sebanyak 6 kali, dan rata-rata *monetary* sebanyak 1167493 rupiah. Namun jumlah pelanggan pada *cluster* 2 lebih sedikit 2215 pelanggan daripada *cluster* 1 sebanyak 2492 pelanggan.

2. Analisis 3 *Cluster*

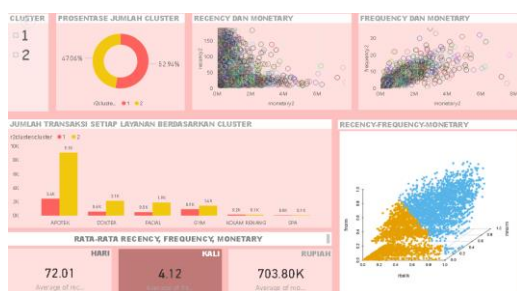
Tabel 12. Hasil Rata-rata 3 *Cluster*

Ckuster	Recency	Frequency	Monetary	Rank
1	23	7	1292044	1
2	80	4	688294	2
3	110	1	165336	3

Berdasarkan hasil rata-rata pada Tabel 12, pada 3 *cluster*, *cluster* 1 menempati peringkat pertama dengan rata-rata *recency*

sebesar 23 hari atau transaksi terakhir yakni pada bulan Oktober, rata-rata *frequency* sebanyak 7 kali, dan rata-rata *monetary* sebanyak 1292044 rupiah, sedangkan *cluster 2* menempati peringkat kedua dengan rata-rata *recency* sebesar 80 hari atau transaksi terakhir yakni pada bulan Agustus, rata-rata *frequency* sebanyak 4 kali, dan rata-rata *monetary* sebanyak 688294 rupiah. *Cluster 3* menempati peringkat ketiga dengan rata-rata *recency* sebesar 110 hari atau transaksi terakhir yakni pada bulan Juli, rata-rata *frequency* sebanyak 1 kali, dan rata-rata *monetary* sebanyak 165336 rupiah. Namun jumlah pelanggan pada *cluster 2* adalah terbanyak dengan jumlah 1617 pelanggan diikuti *cluster 3* sejumlah 1591 pelanggan, dan *cluster 1* sebanyak 1499.

3.8 Hasil Visualisasi Dashboard



Gambar 5. Hasil Visualisasi Halaman Cluster

Berdasarkan Gambar 5 yang merupakan hasil visualisasi penelitian menunjukkan bahwa untuk memvisualisasikan keadaan tiap segmen digunakan beberapa grafik diantaranya grafik *donut* untuk merepresentasikan prosentase pelanggan tiap segmen, grafik *scatter* untuk merepresentasikan hubungan variabel RFM, grafik *bar chart* untuk menampilkan jumlah transaksi layanan setiap *cluster*, label rata-rata variabel RFM, dan *box scatter plot* untuk merepresentasikan plot pelanggan pada ketiga variabel.

3.9 Hasil Usability Testing

Analisis *usability* digunakan untuk mengetahui kualitas *dashboard* yang dibuat. Metode yang digunakan dalam pengukuran *usability* yakni *System Usability Scale* (SUS). Kelebihan metode ini adalah dapat digunakan pada sampel yang jumlahnya sedikit (Sauro, 2011). Dari *testing* yang dilakukan pemilik Belle Crown Malang, didapatkan hasil perhitungan yang dapat disimpulkan bahwa nilai rata-rata

responden untuk *dashboard* adalah 85, sehingga *dashsboard* dikategorikan ke dalam *dashboard acceptable* yang artinya *dashboard* diterima baik oleh perusahaan dalam proses pengambilan keputusan strategis.

4. KESIMPULAN

Berdasarkan tahapan penelitian yang dilakukan, dapat diketahui bahwa:

1. Hasil dari model RFM yakni variabel *Recency*, *Frequency*, dan *Monetary* dapat mendeskripsikan profil atau karakteristik setiap segmen. Hal tersebut dapat dibuktikan dengan nilai *Recency* untuk mengetahui rentang waktu transaksi terakhir pelanggan, nilai *Frequency* untuk mengetahui jumlah transaksi pelanggan, dan *Monetary* untuk mengetahui besaran rupiah yang dikeluarkan pelanggan untuk mendapatkan layanan di Belle Crown Malang.
2. Metode K-Means dapat menjadi suatu pilihan dalam menyelesaikan masalah segmentasi pelanggan. Akan tetapi, hasil pencarian nilai k dengan metode *Elbow* yang telah dilakukan secara *intuitive* sebelum proses *clustering* belum sesuai dengan hasil uji performa yakni nilai jumlah *cluster* pada metode *Elbow* sebanyak 3, sedangkan hasil uji performa sebanyak 2 dan 5. Dari perbedaan hasil tersebut, maka dilakukan *clustering* ulang untuk *cluster* sebanyak 2.
3. Hasil uji performa menggunakan package CValid() pada R yang terdiri dari uji *Dunn Index*, *Connectivity*, dan *Silhouette*. Dari hasil uji performa didapatkan hasil *cluster* yang paling optimal berada pada 2 *cluster* seperti yang terdapat pada hasil uji *Connectivity* dan *Silhouette*.
4. Hasil perhitungan responden dengan metode *System Usability Testing* (SUS) menunjukkan nilai rata-rata yang tinggi yaitu 85 dengan kategori *dashboard acceptable*. Visualisasi ditampilkan secara interaktif yaitu dengan menggabungkan beberapa grafik yang dapat disesuaikan dengan pilihan jenis segmen. Visualisasi grafik *pie chart* menunjukkan prosentase pelanggan setiap layanan. Grafik garis menunjukkan pergerakan jumlah pelanggan pada setiap layanan dari bulan Mei sampaidengan Oktober yang mengalami kenaikan dan penurunan.

Sedangkan grafik *donut chart* menunjukkan prosentase jumlah pembagian pelanggan setiap segmennya, persebaran pelanggan berdasarkan hubungan dari variable *recency-monetary* dan *frequency-monetary* dapat dilihat pada grafik *scatter*, grafik batang merepresentasikan jumlah pelanggan setiap layanan berdasarkan segmennya, karakter setiap segmen yang ditunjukkan melalui nilai rata-rata RFM, persebaran pelanggan berdasarkan variable RFM divisualisasikan melalui *box plot*, melihat jumlah pelanggan berdasarkan *frequency* diinterpretasikan melalui grafik batang dan yang terakhir dapat melihat peringkat pelanggan terbaik dan terburuk setiap segmen.

5. DAFTAR PUSTAKA

- Lestari, Ade Hapsari. 2017. 2020, Industri Kecantikan RI Siap Libas Asia Tenggara. <http://ekonomi.metrotvnews.com/mikro/8N0eqoEb-2020-industri-kecantikan-ri-siap-libas-asia-tenggara>. 15 September 2018 (19:49).
- Aggelis, Vasilis dan Christodoulakis, 2005. "Customer Clustering using RFM Analysis", Proceedings of the 9th WSEAS International Conference on Computers. (ICCOMP'05).
- Angelie, Anissa Veronica, 2017. SEGMENTASI PELANGGAN MENGGUNAKAN CLUSTERING K-MEANS DAN MODEL RFM (STUDI KASUS: PT. BINA ADIDAYA SURABAYA). Institut Teknologi Sepuluh Nopember Surabaya.
- Arthur M. Hughes, 1994. *Strategic database marketing*, Chicago: Probus Publishing.
- Bhatia, S.K., 2012. *A Propound Method For The Improvement of Cluster Quality*. IJCSI International Journal of Computer Science Issues, Vol. 9, issue 4, no 2, 216-221.
- Bholowalia, Purnima & Kumar, Arvind, 2014. EBK-Means: A Clustering Techniques based on Elbow Method and K-Means in WSN. International Journal of Computer Application (0975-8887), IX (105), pp. 17-24.
- Calinski, T., & Harabasz, J., 1974. "A dendrite method for cluster analysis". Communications in Statistics-theory and Methods 3: 1-27.
- Cheng, Ching-Hsue dan Chen, You-Shyang, 2009. "Classifying the segmentation of customer value via RFM model and RS Theory", *Expert Systems with Applications*, Vol.36, hal. 4176-4184.
- G. Brock, V. Pihur, S. Datta dan S. Datta, "clValid, an R package for cluster validation," *Journal of Statistical Software*, 2011.
- Gaffar F., Vanessa, 2007. CRM dan MPR Hotel. Bandung: Alfabeta.
- H. Junaedi, H. Budianto, I. Maryati dan Y. Melani, "Data Transformation Pada Data Mining," Prosiding Konferensi Nasional "Inovasi dalam Desain dan Teknologi", pp. 93-99, 2011.
- Han, Jiawei dan Kamber, Micheline, 2006. *Data Mining : Concept and. Techniques Second Edition*, Morgan Kaufmann Publishers.
- K. Tsipitsis and A. Chorianopoulos, 2009. *Data mining techniques in CRM: inside customer segmentation*. Chichester, West Sussex, U.K: Wiley.
- Kotler, Philip dan Keller, 2007. Manajemen Pemasaran, Jilid I, Edisi. Kedua belas, PT. Indeks, Jakarta.
- M. Khajvand dan M. J. Tarokh, 2011. "Estimating Customer Future Value of Different Customer Segments Based on Adapted RFM Model in Retail Banking Context," *Procedia Computer Science*, vol. 3, p. 1327-1332.
- P. N. Tan, M. Steinbach dan V. Kumar, 2005. *Introduction to Data Mining*, Boston: Addison-Wesley Longman Publishing Co.
- Rainer and Cegielski, 2011. *Introduction to Information Systems. (3rd Edition)*. USA: Wiley.

- Sauro, Jeff, 2011. "Measuring Usability With The System Usability Scale (SUS)", <http://www.measuringu.com/sus.php>. Diakses pada 17 Januari 2018.
- Sauro, Jeff, 2013. " A Single-Item Measure of Website Usability: Comments on Christophersen and Konradt (2011) Interacting with Computers, Special Issue: Commentary on Scale Derivation", 25 (4): 302303.
- Y.-S. C. Ching-Hsue Cheng, 2009. "*Classifying the segmentation of customer value via RFM model and RS theory*," *Expert system with application*, vol. 36, pp. 4176-4184.