

Implementasi *Fuzzy C-Means* dan Model RFM untuk Segmentasi Pelanggan (Studi Kasus : PT. XYZ)

Denny B. Saputra dan Edwin Riksakomara

Departemen Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember (ITS)

e-mail: riksakomara@gmail.com

Abstrak—Era globalisasi yang disertai dengan berlakunya MEA menyebabkan persaingan dalam berbagai sektor bisnis, termasuk sektor industri. Hal ini memaksa perusahaan-perusahaan yang bergerak di sektor industri di Indonesia untuk bersaing dengan negara-negara tetangga di ASEAN, termasuk PT. XYZ. Banyaknya perusahaan kompetitor yang menawarkan harga rendah membuat pelanggan PT. XYZ beralih pada kompetitor. Terkait dengan hal tersebut, PT. XYZ telah melakukan berbagai kegiatan promosi dalam rangka mengelola hubungan dengan pelanggan. Namun kegiatan promosi selama ini diimplementasikan kepada semua pelanggan tanpa adanya segmentasi sehingga strategi promosi menjadi tidak efektif dan efisien. Untuk menyelesaikan masalah pelanggan tersebut, salah satu solusi yang dapat dilakukan adalah menentukan segmentasi pelanggan. Segmentasi yang dilakukan menggunakan metode clustering *Fuzzy C-Means*. Sedangkan untuk menentukan jumlah K , metode yang digunakan adalah metode *Elbow*. Dalam proses clustering, diterapkan variabel RFM yang menggambarkan perilaku pelanggan dalam melakukan transaksi. Setelah hasil clustering muncul, cluster tersebut akan divisualisasikan sehingga memudahkan untuk menganalisa setiap cluster. Hasil dari clustering dengan *Fuzzy C-Means* dan model RFM adalah tiga segmen pelanggan. Segmen 1 adalah segmen terburuk dengan rata-rata nilai CLV 0.0538. Segmen 2 adalah segmen menengah dengan rata-rata nilai CLV 0.1674. Segmen 3 adalah segmen terbaik dengan rata-rata nilai CLV 0.3516. Segmen 1 memiliki 42 anggota, segmen 2 memiliki 64 anggota, dan segmen 3 memiliki 17 anggota

Kata Kunci—*Clustering, Fuzzy C-Means, Model RFM, Segmentasi Pelanggan.*

I. PENDAHULUAN

ERA globalisasi yang beberapa tahun lalu dimulai, menyebabkan perubahan besar bagi banyak negara. Terlebih setelah disepakatinya MEA pada awal tahun 2016, menyebabkan dampak buruk pada berbagai sektor, termasuk sektor industri [1]. Hal ini memaksa perusahaan-perusahaan yang bergerak di sektor industri di Indonesia untuk bersaing dengan negara-negara tetangga di ASEAN. Salah satu industri yang saat ini tengah bersaing adalah industri pertambangan dan minyak bumi. Industri pertambangan dan penggalian mencatat penurunan pertumbuhan di tengah mengilapnya perekonomian nasional [2].

PT. XYZ merupakan cabang perusahaan yang bergerak di industri tambang dan minyak bumi. Seperti halnya pertumbuhan industri tambang yang melemah di kuartal I tahun 2017, PT. XYZ juga mengalami hal serupa. Selama bulan Januari hingga Juni 2017, angka penjualan produk tersebut tidak mencapai target yang ditetapkan setiap bulannya [3]. Adapun tidak tercapainya target penjualan disebabkan berbagai faktor. Salah satunya adalah harga jual produk lebih rendah yang ditawarkan oleh kompetitor menyebabkan sebagian pelanggan beralih kepada kompetitor. Terkait dengan masalah ini, PT. XYZ telah melakukan berbagai kegiatan promosi untuk menarik kembali pelanggan [3]. PT. XYZ menerapkan strategi marketing kepada pelanggan tanpa adanya segmentasi pelanggan. Oleh karena itu, PT. XYZ menerapkan strategi yang sama untuk semua pelanggan. Penerapan strategi yang sama ini dapat menimbulkan beberapa dampak, yaitu strategi promosi yang tidak tepat sasaran sehingga strategi tersebut hanya bermanfaat pada pelanggan tertentu saja. Selain itu, waktu dan dana yang digunakan untuk promosi menjadi kurang efektif dan efisien.

Untuk menyelesaikan masalah pelanggan tersebut, salah satu solusi yang dapat dilakukan adalah menentukan segmentasi pelanggan. Banyak metode yang dapat digunakan dalam melakukan segmentasi pelanggan, salah satunya adalah penggalan data terhadap riwayat transaksi pembelian pelanggan pada periode tertentu. Data transaksi tersebut kemudian ditentukan variabel-variabel RFM. Data transaksi yang digunakan adalah periode Januari 2017 hingga Juni 2017. Algoritma *clustering* yang digunakan adalah *Fuzzy C-Means*. *Fuzzy C-Means* merupakan salah satu teknik untuk menentukan *cluster* optimal pada suatu ruang vektor yang didasarkan pada bentuk normal *Euclidean* untuk jarak antar vektor. *Fuzzy C-Means* menggunakan derajat keanggotaan sebagai dasar penentuan suatu vektor masuk ke dalam *cluster* tertentu. Untuk mengecek validitas hasil *clustering*, digunakan *Sum of Square Error (SSE)* sebagai metode uji. Semakin kecil nilai SSE, maka semakin bagus hasil *clustering* yang dilakukan.

Dengan hasil penelitian ini, diharapkan dapat memberikan pendukung dalam menentukan strategi pemasaran terhadap pelanggan, dengan bantuan visualisasi berbasis *web* sehingga mempermudah analisis.

II. TINJAUAN PUSTAKA

A. Segmentasi Pelanggan

Segmentasi pelanggan adalah dasar bagaimana memaksimalkan nilai dari pelanggan [4]. Segmentasi dapat dipandang sebagai kompleksitas yang berhubungan dengan pelanggan secara individual, masing-masing dengan kebutuhan dan nilai potensial yang berbeda [5]. Segmentasi pelanggan menghasilkan segmen berdasarkan karakteristik tertentu. Karakteristik yang dapat menjadi acuan antara lain [6]:

- a. Demographics : umur, jenis kelamin, besarnya keluarga, besarnya tempat tinggal, siklus kehidupan keluarga, pemasukan, pekerjaan, pendidikan kepemilikan rumah, status sosial ekonomi, agama, kewarganegaraan.
- b. Psychographics : kepribadian, gaya hidup, nilai-nilai, sikap
- c. Behavior : manfaat yang dicari, status pembelian, tingkat penggunaan produk, frekuensi pembelian
- d. Geographics : negara, provinsi, kota, kode pos, iklim

B. Model RFM

Untuk mengidentifikasi kebiasaan pelanggan, metode yang umum digunakan adalah *recency, frequency, and monetary* (RFM) model yang digunakan untuk merepresentasikan karakteristik kebiasaan pelanggan [7]. Model RFM pertama kali diperkenalkan oleh Hughes dan saat ini banyak digunakan oleh industri termasuk manufaktur, *retailer*, dan industri jasa [8]. Model RFM melibatkan 3 variabel yaitu *Recency of the last purchases, Frequency of the purchases*, dan *Monetary value of the purchases* [9]. Adapun penjelasan ketiga variabel tersebut adalah:

- a. *Recency* adalah jarak dari waktu transaksi terakhir kali dilakukan dengan waktu saat ini. Semakin kecil nilai jarak waktu, maka semakin besar nilai R.
- b. *Frequency* adalah total jumlah transaksi yang dilakukan selama periode tertentu. Semakin besar jumlah transaksi, maka semakin besar nilai F.
- c. *Monetary* adalah total nilai produk dalam bentuk uang dalam periode tertentu. Semakin besar nilai produk, maka semakin besar nilai M.

C. Metode Fuzzy C-Means

Fuzzy C-Means (FCM) pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. FCM merupakan salah satu teknik untuk menentukan *cluster* optimal pada suatu ruang vektor yang didasarkan pada bentuk normal *Euclidean* untuk jarak antar vektor. FCM akan meng*cluster* data dimana posisi tiap-tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Algoritma *Fuzzy C-Means* adalah sebagai berikut [10]:

a. Input Data

Input data yang akan *dicluster* X, berupa matriks berukuran $n \times m$, dengan n = jumlah sampel data, m = atribut setiap data). X_{ij} = data sampel ke- i ($i = 1,2,3,\dots,n$), atribut j ($j = 1,2,3,\dots,m$)

b. Menentukan Nilai Variabel

- Jumlah cluster = c ;
- Pangkat/pembobot = w ;
- Maksimum Iterasi = $MaxIter$;
- Error terkecil yang diharapkan = ξ ;
- Fungsi objektif awal = $P_0 = 0$;
- Iterasi awal = $t = 1$;

c. Membangkitkan Nilai Random

Bangkitkan nilai random μ_{ik} , $i = 1,2,3,\dots,n$; $k = 1,2,3,\dots,c$ sebagai elemen – elemen matriks partisi awal μ_{ik} . μ_{ik} adalah derajat keanggotaan yang merujuk pada seberapa besar kemungkinan suatu data bisa menjadi anggota ke dalam suatu *cluster*. Hitung jumlah setiap kolom (atribut) :

$$Q_j = \sum_{k=1}^c \mu_{ik}, j = 1,2,3,\dots,m \tag{1}$$

Hitung :

$$\mu_{ik} = \frac{\mu_{ik}}{Q_j} \tag{2}$$

d. Menghitung Pusat Cluster ke-k

Hitung pusat cluster ke- k : V_{kj} dengan $k = 1,2,3,\dots,c$; $j = 1,2,3,\dots,m$.

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * x_{ij})}{\sum_{i=1}^n ((\mu_{ik})^w)} \tag{3}$$

e. Menghitung Fungsi Objektif

Hitung fungsi objektif pada iteasi ke- t , P_t :

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (x_{ij} - v_{kj})^2 \right] (u_{ik})^w \right) \tag{4}$$

f. Menghitung Perubahan Matriks

Hitung perubahan matriks partisi :

$$u_{ik} = \frac{\left[\sum_{j=1}^m (x_{ij} - v_{kj})^2 \right]^{-\frac{1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (x_{ij} - v_{kj})^2 \right]^{-\frac{1}{w-1}}}; i = 1,2,3,\dots,n; k = 1,2,3,\dots,c \tag{5}$$

g. Mengecek Kondisi Berhenti

- Jika $(|P_t - P_{t-1}| < \xi)$ atau $(t > MaxIter)$ maka berhenti;
- Jika tidak, $t = t + 1$, ulangi langkah 4.

D. Pengujian SSE

Sum Square Error (SSE) adalah salah satu formula sederhana yang digunakan untuk mengevaluasi hasil *cluster*. Dalam perhitungan SSE, setiap objek pada *cluster* akan dihitung tingkat kesalahan terhadap *centroid* terdekat. Formula untuk menghitung nilai SSE adalah [11]:

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} dist(m_1, x)^2 \tag{6}$$

III. METODOLOGI PENELITIAN

Penelitian dilakukan dengan tahapan pelaksanaan seperti dijelaskan pada Gambar 1.

A. Identifikasi Masalah

Tahap ini merupakan tahap awal dalam pengerjaan tugas akhir ini. Pada tahap ini dilakukan analisis permasalahan yang akan diselesaikan pada tugas akhir ini, dimana permasalahan yang diangkat adalah segmentasi pelanggan.

B. Studi Literatur

Pada tahap ini dilakukan pencarian dan pengumpulan berbagai referensi penelitian terdahulu. Literatur yang dibahas adalah terkait dengan penggalian data, *clustering*, *Fuzzy C-Means*, metode RFM, dan segmentasi pelanggan

C. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data-data yang diperlukan untuk penelitian tugas akhir, terutama yang terkait dengan proses *clustering*.

D. Praproses Data

Praproses data dibutuhkan agar data mentah yang telah dikumpulkan dapat diolah sesuai dengan kriteria yang diinginkan. Tahap praproses data ini meliputi *data cleaning*, pemodelan RFM dan normalisasi.

E. Pembobotan RFM

Pada tahapan ini dilakukan pembobotan terhadap variabel RFM dengan metode AHP. Dari perhitungan AHP tersebut, diperoleh nilai bobot untuk variabel *recency*, *frequency*, dan *monetary*.

F. Perhitungan CLV

Perhitungan *CLV* dilakukan dengan perkalian nilai variabel RFM dengan bobot masing-masing. Indeks *CLV* didapatkan dari penjumlahan semua nilai variabel RFM.

G. Clustering Data

Sebelum *clustering* dilakukan, ditentukan terlebih dahulu nilai *K* sebagai jumlah *cluster* yang akan dibentuk. Metode yang digunakan untuk menentukan jumlah *K* adalah *Elbow*. Metode *clustering* yang digunakan adalah *Fuzzy C-Means*.

H. Pengujian Kualitas Cluster

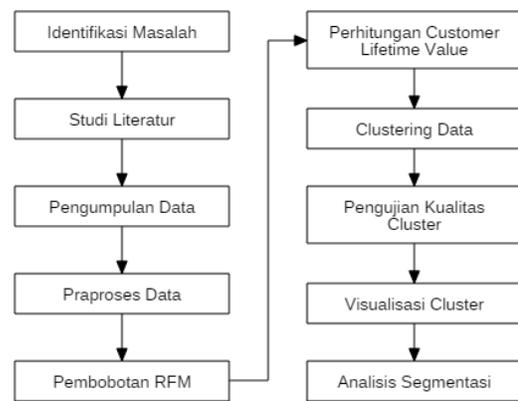
Pengujian hasil *cluster* dilakukan dengan menguji validitas berdasarkan *SSE*. Nilai *SSE* diperoleh dari rata-rata jarak semua objek terhadap titik pusat *cluster* pada masing-masing *cluster*.

I. Visualisasi Cluster

Pada tahap visualisasi *cluster*, hasil *cluster* akan divisualisasikan dengan media dashboard berbasis web dengan alat bantu RShiny. Visualisasi ini meliputi grafik *scattered plot* 3D, *pie chart*, dan *Box and Whisker*.

J. Analisis Segmentasi

Analisis segmentasi dilakukan untuk menjelaskan ciri atau



Gambar. 1. Metodologi Penelitian

kemiripan anggota dalam satu segmen, serta perbandingan dengan segmen lainnya. Analisis ini juga bertujuan untuk mengidentifikasi segmen mana yang menjadi prioritas dibanding segmen yang lain.

IV. HASIL DAN PEMBAHASAN

A. Pembobotan RFM dengan AHP

Data kuesioner diolah dengan cara merata-rata semua pengisian responden terhadap perbandingan tiga kriteria RFM. Hasil rata-rata ini dijumlahkan setiap kolomnya. Setelah itu dilakukan normalisasi matriks pada hasil rata-rata nilai setiap kriteria. Tabel baru yang dihasilkan tersebut disebut sebagai *normal comparisons matrix*. Hasil pembobotan dapat dilihat pada Tabel 1.

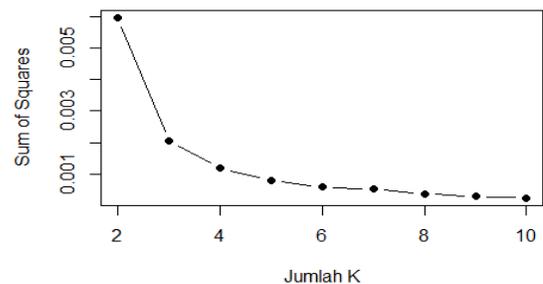
Tabel 1.

Hasil Pembobotan Variabel RFM

Kriteria	Recency	Frequency	Monetary	Bobot
Recency	1	2/7	1/5	0.10
Frequency	3 2/3	1	4/9	0.31
Monetary	5	2 1/3	1	0.59

B. Penentuan Nilai K

Penentuan nilai *K* dilakukan menggunakan metode *Elbow*. Hasil perhitungan metode *Elbow* ditampilkan dalam plot dua dimensi pada Gambar 2. Nilai pada sumbu X adalah jumlah *K* yaitu 2 hingga 10, sedangkan sumbu Y adalah nilai *SSE* yang dihasilkan oleh setiap nilai *K*. Metode *Elbow* menjelaskan



Gambar. 2. Penentuan Nilai K

bahwa nilai *K* yang diambil adalah pada titik dimana terjadi penurunan yang signifikan dan diikuti oleh nilai yang relatif konstan. Dapat dilihat bahwa titik yang dimaksud adalah 3 atau

4. Akan tetapi penurunan pada $K = 3$ lebih signifikan dibanding dengan $K = 4$, sehingga nilai 3 dipilih sebagai nilai K untuk proses *clustering*.

C. *Clustering Fuzzy C-Means*

Proses *Clustering Fuzzy C-Means* dilakukan pada RStudio. Luaran yang dihasilkan dari eksekusi script adalah pusat setiap *cluster*, derajat keanggotaan setiap objek terhadap tiap *cluster*, dan *cluster* tiap objek berdasarkan derajat keanggotaannya yang paling besar terhadap *cluster* tertentu.

D. *Pengujian SSE*

Pengujian SSE dilakukan dengan menggunakan aplikasi

```
Iteration: 10, Error: 0.0020572052
Iteration: 11, Error: 0.0020514517
Iteration: 12, Error: 0.0020490784
Iteration: 13, Error: 0.0020481295
Iteration: 14, Error: 0.0020477582
Iteration: 15, Error: 0.0020476149
Iteration: 16, Error: 0.0020475602
Iteration: 17, Error: 0.0020475393
Iteration: 18, Error: 0.0020475314
Iteration: 19, Error: 0.0020475285
Iteration: 20, Error: 0.0020475273
Iteration: 21, Error: 0.0020475269
Iteration: 22, Error: 0.0020475267
Iteration: 23, Error: 0.0020475267
Iteration: 24 converged, Error: 0.0020475267
> print(c1$withinerror)
[1] 0.002047527
```

Gambar. 4. Hasil Pengujian SSE

RStudio. Hasil pengujian SSE menampilkan nilai *error* pada hasil *clustering* berdasarkan parameter yang telah digunakan. Nilai SSE yang diperoleh untuk keseluruhan data adalah 0.0020475. Hasil pengujian SSE dapat dilihat pada Gambar 4.

E. *Analisis Clustering*

Tabel 2. Perbandingan Karakteristik Antar Segmen

	Segmen 1	Segmen 2	Segmen 3
<i>Peringkat</i>	3	2	1
<i>Jumlah Anggota</i>	42	64	17
<i>Recency (hari)</i>	110 - 256	93 - 122	93 - 110
<i>Frequency (kali)</i>	1 - 6	2 - 25	8 - 60
<i>Monetary (Rp)</i>	26.800 - 51.101.240	133.893 - 816.972.362	7.481.951 - 861.123.629

Segmen 1 merupakan segmen dengan peringkat ketiga atau peringkat terendah berdasarkan rata-rata nilai CLV. Segmen 1 yang beranggotakan 42 pelanggan memiliki nilai *recency* sangat besar, yang menunjukkan bahwa pelanggan sudah lama tidak aktif bertransaksi. Nilai *frequency* yang rendah menunjukkan bahwa pelanggan jarang melakukan transaksi. Selain itu, nilai *monetary* juga rendah, yang menunjukkan bahwa pelanggan mengeluarkan sedikit uang saat bertransaksi.

Segmen 2 merupakan segmen dengan peringkat kedua berdasarkan rata-rata nilai CLV. Segmen 2 yang beranggotakan 64 pelanggan memiliki nilai *recency* lebih kecil, yang menunjukkan bahwa pelanggan aktif bertransaksi dalam waktu dekat. Nilai *frequency* yang lebih besar menunjukkan bahwa

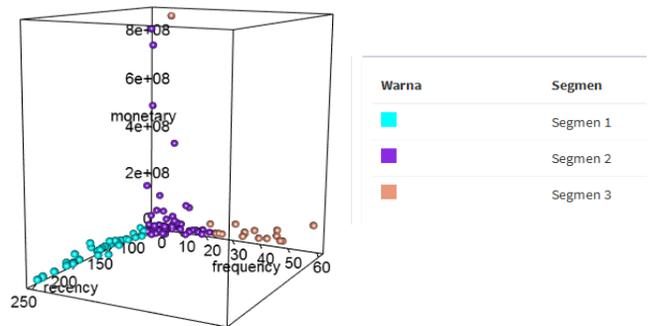
pelanggan cukup sering melakukan transaksi. Nilai *monetary* juga lebih besar, yang menunjukkan pelanggan mengeluarkan lebih banyak uang saat bertransaksi.

Segmen 3 merupakan segmen dengan peringkat pertama atau segmen terbaik berdasarkan rata-rata nilai CLV. Segmen 3 yang beranggotakan 17 pelanggan memiliki nilai *recency* paling kecil, yang menunjukkan bahwa pelanggan aktif bertransaksi dalam waktu dekat. Nilai *frequency* yang paling besar menunjukkan bahwa pelanggan sangat sering melakukan transaksi. Selain itu, nilai *monetary* juga paling besar, yang menunjukkan bahwa pelanggan mengeluarkan paling banyak uang ketika bertransaksi.

F. *Analisis Visualisasi*

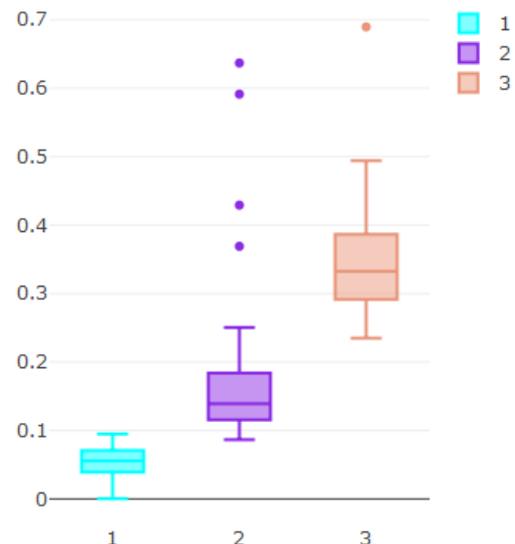
1) *Scatter Plot 3D*

Dari Gambar 4 dapat diketahui bahwa pelanggan pada segmen 1 cenderung berkumpul pada variabel *recency* dan mengarah pada nilai *recency* yang semakin tinggi. Pelanggan pada segmen 2 cenderung berkumpul mendekati titik 0 dan mengarah pada variabel *monetary* yang semakin tinggi. Pelanggan pada segmen 3 cenderung berkumpul pada variabel *frequency* dan mengarah ke nilai *frequency* yang semakin tinggi.



Gambar 4. Grafik Box Plot Nilai Recency

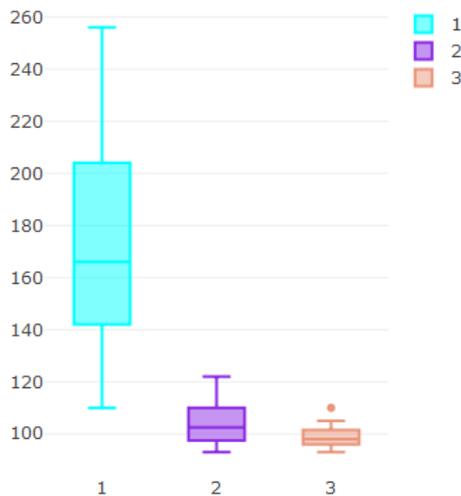
2) *Box Plot CLV*



Gambar 5. Grafik Box Plot Nilai CLV

Segmen 1 memiliki ukuran kotak yang paling pendek dibanding dengan dua segmen lain. Hal ini menunjukkan bahwa sebaran data pada segmen 1 juga kecil. Nilai median segmen 1 terletak di tengah kotak dan garis *whisker* sama panjang, yang berarti segmen 1 bersifat simetris. Garis *whisker* bagian atas sedikit berurusan dengan segmen 2, sehingga menyebabkan sedikit *overlap*. Segmen 2 memiliki kotak yang lebih panjang, sehingga sebaran data pada segmen 2 lebih lebar. Nilai median segmen 2 tidak terletak di tengah kotak dan panjang garis *whisker* bagian atas lebih panjang. Hal ini menunjukkan bahwa segmen 2 tidak bersifat simetris dan cenderung ke *positive skewness*. Selain itu terdapat beberapa data yang *outlier*, sehingga *overlap* terhadap segmen 3. Segmen 3 memiliki kotak yang paling panjang, sehingga persebaran data menjadi lebar. Nilai median segmen 3 tidak terletak di tengah kotak dan panjang garis *whisker* bagian atas lebih panjang. Hal ini menunjukkan bahwa segmen 3 tidak simetris dan cenderung ke *positive skewness*. Selain itu terdapat data *outlier* di bagian atas kotak.

3) *Box Plot Recency*



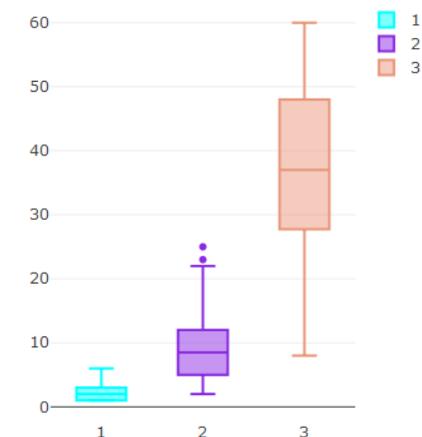
Gambar 6. Grafik Box Plot Nilai Recency

Segmen 1 memiliki ukuran kotak yang paling panjang dibanding dengan dua segmen lain, sehingga persebaran data menjadi lebar. Nilai median segmen 1 tidak berada di tengah kotak dan garis *whisker* bagian atas lebih panjang. Hal ini menunjukkan menunjukkan bahwa segmen 2 tidak bersifat simetris dan cenderung ke *positive skewness*. Garis *whisker* bagian bawah sedikit berurusan dengan segmen 2 sehingga menyebabkan sedikit *overlap*. Segmen 2 memiliki kotak lebih pendek, sehingga persebaran data lebih kecil dibanding segmen 1. Nilai median segmen 2 tidak terletak di tengah kotak dan panjang garis *whisker* bagian atas lebih panjang. Hal ini menunjukkan bahwa segmen 2 tidak bersifat simetris dan cenderung ke *positive skewness*. Sebagian kotak segmen 2 berurusan dengan segmen 3, sehingga menyebabkan banyak *overlap*. Segmen 3 memiliki kotak yang paling pendek, sehingga persebaran data menjadi paling kecil. Nilai median segmen 3 tidak terletak di tengah kotak dan panjang garis *whisker* bagian atas lebih panjang. Hal ini menunjukkan bahwa

segmen 3 tidak simetris dan cenderung ke *positive skewness*. Selain itu terdapat data *outlier* di bagian atas kotak.

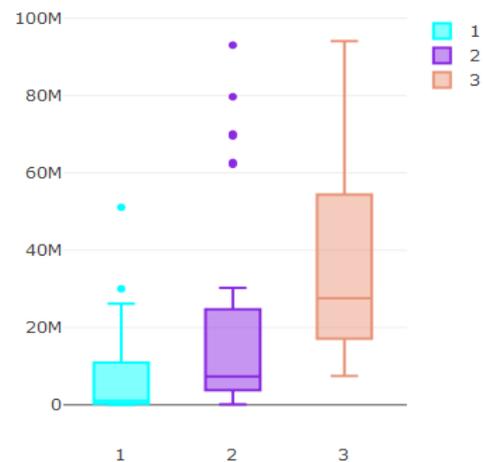
4) *Box Plot Frequency*

Segmen 1 memiliki ukuran kotak yang paling pendek dibanding dengan dua segmen lain. Hal ini menunjukkan bahwa sebaran data pada segmen 1 juga kecil. Nilai median segmen 1 terletak di tengah kotak, tetapi garis *whisker* bagian atas lebih panjang. Sebagian kotak berurusan dengan garis *whisker* segmen 2, sehingga terjadi sedikit *overlap*. Segmen 2 memiliki kotak yang lebih panjang, sehingga sebaran data pada segmen 2 lebih lebar. Nilai median segmen 2 terletak di tengah kotak, tetapi garis *whisker* bagian atas lebih panjang. Selain itu terdapat beberapa data yang *outlier* di bagian atas kotak. Sebagian kotak berurusan dengan garis *whisker* segmen 3, sehingga terjadi sedikit *overlap*. Segmen 3 memiliki kotak yang paling panjang, sehingga persebaran data menjadi paling lebar. Nilai median segmen 3 tidak terletak di tengah kotak dan panjang garis *whisker* bagian bawah lebih panjang. Hal ini menunjukkan bahwa segmen 3 tidak simetris dan cenderung ke *negative skewness*.



Gambar 7. Grafik Box Plot Nilai Frequency

5) *Box Plot Monetary*



Gambar 8. Grafik Box Plot Nilai Monetary

Segmen 1 memiliki ukuran kotak yang paling pendek dibanding dengan dua segmen lain. Hal ini menunjukkan bahwa sebaran data pada segmen 1 juga paling kecil. Nilai median segmen 1 tidak terletak di tengah kotak dan garis *whisker* bagian atas lebih panjang. Hal ini menunjukkan menunjukkan bahwa segmen 1 tidak bersifat simetris dan cenderung ke *positive skewness*. Selain itu, terdapat beberapa data yang *outlier* di bagian atas kotak. Sebagian kotak beririsan dengan kotak segmen 2 dan segmen 3, sehingga terjadi banyak *overlap*. Segmen 2 memiliki kotak yang lebih panjang, sehingga sebaran data pada segmen 2 lebih lebar. Nilai median segmen 2 tidak terletak di tengah kotak dan garis *whisker* bagian atas lebih panjang. Hal ini menunjukkan menunjukkan bahwa segmen 2 tidak bersifat simetris dan cenderung ke *positive skewness*. Selain itu terdapat beberapa data yang *outlier* di bagian atas kotak. Sebagian kotak beririsan dengan kotak segmen 1 dan 3, sehingga terjadi banyak *overlap*. Segmen 3 memiliki kotak yang paling panjang, sehingga persebaran data menjadi paling lebar. Nilai median segmen 3 tidak terletak di tengah kotak dan panjang garis *whisker* bagian atas lebih panjang. Hal ini menunjukkan bahwa segmen 3 tidak simetris dan cenderung ke *positive skewness*. Sebagian kotak beririsan dengan kotak segmen 2 sehingga terjadi banyak *overlap*.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

- Hasil AHP menunjukkan bahwa variabel *monetary* memiliki bobot tertinggi yaitu sebesar 0.59. Hal ini berarti bahwa variabel *monetary* menjadi pertimbangan paling penting dalam melakukan *clustering*. Sedangkan variabel *recency* dan *monetary* masing-masing memiliki bobot 0.31 dan 0.10.
- Berdasarkan metode Elbow, nilai K yang paling baik untuk digunakan adalah 3, sehingga jumlah *cluster* yang dihasilkan pada proses *clustering* berjumlah 3 *cluster*.
- Dari 123 pelanggan yang di-*cluster*, 42 pelanggan tergabung pada *cluster* 1, 64 pelanggan tergabung pada *cluster* 2, dan 17 pelanggan tergabung pada *cluster* 3.
- Cluster* 1 mempunyai rata-rata nilai CLV paling rendah. Hal ini didukung dengan nilai *recency* yang sangat tinggi yang mengindikasikan bahwa pelanggan sudah lama tidak melakukan transaksi. Sehingga dapat disimpulkan bahwa *cluster* 1 adalah *cluster* terburuk.
- Cluster* 2 memiliki rata-rata CLV sedang, begitu juga dengan nilai ketiga variabel. Sehingga dapat disimpulkan bahwa *cluster* 2 adalah *cluster* urutan kedua.
- Cluster* 3 memiliki rata-rata CLV tertinggi. Hal ini didukung dengan nilai *frequency* dan *monetary* yang tinggi yang mengindikasikan pelanggan sering melakukan transaksi dan mengeluarkan banyak uang. Sehingga dapat disimpulkan bahwa *cluster* 3 adalah *cluster* terbaik.

B. Saran

- Penelitian ini dapat dikembangkan dengan memperhatikan atribut lain selain atribut RFM untuk memberikan variasi hasil *clustering*, misalkan dengan menggunakan lokasi pelanggan.
- Rentang data transaksi sebagai masukan lebih baik diperbanyak lebih dari 6 bulan transaksi agar hasil *clustering* lebih akurat, misalkan selama 1 tahun.
- Penelitian ini dapat dilanjutkan dengan menggunakan metode *clustering* yang lain selain *Fuzzy C-Means* atau dikombinasikan dengan metode lain, misalkan ANFIS.
- Penentuan nilai K dapat dikembangkan dengan metode lain, misalkan SOM, agar memberikan nilai K yang lebih optimal.

DAFTAR PUSTAKA

- Tempo.co, "Inilah Industri Yang Rawan Kena Dampak MEA," www.tempo.co, 2016. .
- CNN Indonesia, "Ekonomi Tumbuh, Industri Tambang Malah Melorot," 2017. .
- Petrochemical Trading, "Monthly Report Petrochemical Trading PT. XYZ Bulan Januari s/d Juni 2017," 2017.
- Y. Chen, G. Zhang, D. Hu, and S. Wang, "Knowledge Enterprise: Intelligent Strategies In Product Design, Manufacturing, and Management," in *International Federation for Information Processing (IFIP)*, 2006, pp. 288–293.
- W. Hugh, D. Elizabeth, and M. Malcolm, "Factors for Success in Customer Relationship Management (CRM) Systems," *J. Mark. Manag.*, vol. 18, pp. 193–219, 2002.
- M. J. A. Berry and G. S. Linoff, *Mastering Data Mining The Art and Science of Customer Relationship Management*. Canada: John Wiley & Sons Inc, 2000.
- C. Chan, "Online Auction Customer Segmentation Using A Neural Network Model," *Int. J. Appl. Sci. Eng.*, vol. 3, pp. 101–109, 2005.
- Y.-S. Chen, C.-H. Cheng, C.-J. Lai, C.-Y. Hsu, and H.-J. Syu, "Identifying patients in target customer segments using a two-stage clustering-classification approach: A hospital-based assessment," *Comput. Biol. Med.*, pp. 213–221, 2012.
- A.M. Hughes, *Strategic Database Marketing*. Chicago: Probus Publishing Company, 2012.
- Kusumadewi, *Aplikasi Logika Fuzzy untuk pendukung keputusan*. Yogyakarta: Graha Ilmu, 2013.
- P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Boston: Addison-Wesley Longman Publishing Co, 2005.