

PENERAPAN *FUZZY C-MEANS* KLUSTER UNTUK SEGMENTASI PELANGGAN *E-COMMERCE* DENGAN METODE *RECENCY FREQUENCY* *MONETARY* (RFM)

Stevanus Sandy Prasetyo¹, Mustafid², Arief Rachman Hakim³

^{1,2,3}Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

E-mail: sandi.prasetya09.sp@gmail.com

ABSTRACT

E-commerce has become a medium for online shopping which is growing and popular among the public. Due to the ease of access for all internet users and the completeness of the products offered, e-commerce has become a new alternative in meeting people's needs. Currently, the competition in the business world is very fierce, any e-commerce company needs to be able to carry out the right marketing strategy to compete in acquiring, retaining, and partnering with customers. In this research, the segmentation of e-commerce customers was carried out using the Fuzzy C-Means cluster and the RFM method. The clustering process is carried out six times with the number of clusters starts from two to seven clusters. The results showed that the optimum number of clusters formed according to the Xie-Beni validity index was four clusters. The cluster becomes customer segments that have the characteristics of each customer based on their recency, frequency, and monetary value. The best segment is segment 4 which has very loyal customers in shopping on tumbas.in e-commerce. From the segments that have been formed, they can be used as a consideration in implementing the right marketing strategy for each customer.

Keywords : E-commerce, customer segmentation, Fuzzy C-Means Cluster, RFM, Xie-Beni Index.

1. PENDAHULUAN

Dunia bisnis saat ini berkembang dengan sangat pesat, hal ini sejalan dengan kemajuan teknologi internet yang dapat memfasilitasi segala aktifitas bisnis. Internet telah menjadi sarana penunjang untuk berbelanja pada sebuah *marketplace* baru yang disebut sebagai *e-commerce*. *E-commerce* itu sendiri merupakan *platform* ataupun wadah yang memfasilitasi proses jual beli barang maupun jasa secara *online* (Wong, 2010). Indonesia menjadi salah satu pasar yang menarik perhatian para pengusaha *e-commerce*. Menurut survei pada tahun 2019, tercatat bahwa pengguna internet di Indonesia pada tahun 2018 telah mencapai 171,17 juta pengguna dan tentunya akan bertambah setiap tahunnya (APJII, 2019).

Lonjakan dari pertumbuhan pengguna internet ini tentunya juga berdampak terhadap pertumbuhan jumlah pelanggan pada setiap perusahaan *e-commerce*. Dalam hal ini menjadi penting bagi setiap perusahaan *e-commerce* untuk melakukan strategi pemasaran yang tepat agar dapat bersaing dalam memperoleh, mempertahankan, dan bermitra dengan pelanggan. Salah satu strategi pemasaran yang dapat dilakukan adalah dengan membuat segmentasi pelanggan, segmentasi pelanggan dapat membantu perusahaan dalam mengelompokkan setiap pelanggannya berdasarkan kesamaan karakteristiknya. Namun permasalahan yang masih dihadapi saat ini oleh banyak perusahaan *e-commerce* di Indonesia adalah sulitnya menentukan dasar yang tepat untuk membentuk segmentasi pelanggannya. Untuk menyelesaikan permasalahan tersebut salah satu solusi yang dapat dilakukan adalah dengan membuat segmentasi pelanggan berdasarkan data histori transaksi pelanggan.

Metode *Recency Frequency Monetary* (RFM) dapat digunakan untuk menganalisa data histori transaksi pelanggan, karena metode ini mampu menganalisa karakteristik pelanggan berdasarkan database transaksi pelanggan dengan membentuknya kedalam tiga

variabel utama yaitu *recency*, *frequency*, dan *monetary* (Mohammadian dan Makhani, 2019). Untuk dapat membentuk segmentasi pelanggan guna mengelompokkan pelanggan berdasarkan karakteristiknya, salah satu metode yang dapat digunakan adalah dengan analisis kluster. Terdapat beberapa algoritma analisis kluster salah satunya adalah algoritma Fuzzy C-Means. Metode ini merupakan salah satu teknik untuk menentukan kluster optimal pada suatu ruang vektor yang didasarkan pada bentuk normal euclidean untuk jarak antar vektor, dan menggunakan derajat keanggotaan sebagai dasar penentuan suatu vektor masuk kedalam kluster tertentu (Saputra dan Riksakomara, 2018).

Penelitian ini bertujuan untuk menerapkan algoritma Fuzzy C-Means dan metode *Recency Frequency Monetary* (RFM) dalam membentuk segmen-segmen dari pelanggan *e-commerce* tumbas.in. Penelitian menggunakan data histori transaksi pelanggan dalam kurun waktu pada bulan Maret 2018-Agustus 2018, dan diolah dengan menggunakan perangkat lunak R i386.3.6.2.

2. TINJAUAN PUSTAKA

2.1. Segmentasi Pelanggan

Segmentasi pelanggan adalah proses membagi basis pelanggan menjadi kelompok kelompok homogen yang berbeda, untuk mengembangkan strategi pemasaran yang berbeda sesuai dengan karakteristiknya (Tsiptsis, 2010). Perusahaan mengimplementasikan segmentasi pelanggan berdasarkan fakta bahwa setiap pelanggan berbeda dan usaha pemasaran akan terlayani dengan baik jika terdapat target yang spesifik. Segmentasi umumnya digunakan untuk mendapatkan wawasan tentang sikap pelanggan, keinginan, pandangan, preferensi, opini tentang perusahaan dan kompetisi.

2.2. Metode Recency Frequency Monetary (RFM)

Untuk mengidentifikasi aktifitas transaksi pelanggan berdasarkan data transaksi, metode yang umum digunakan adalah *recency frequency dan monetary* (RFM). Metode RFM membentuk data histori transaksi pelanggan kedalam tiga variabel yaitu *recency of the last purchase*, *frequency of the purchase*, dan *monetary value purchase* (Hughes, 1994). Adapun penjelasan ketiga variabel tersebut adalah :

a. *Recency* adalah jarak dari waktu transaksi terakhir kali dilakukan dengan waktu dilakukannya penelitian. Variabel *recency* diperoleh dari selisih waktu transaksi yang terakhir kali dilakukan oleh masing-masing pelanggan dengan tanggal yang telah ditetapkan untuk penelitian. Nilai *recency* diperoleh dengan menggunakan persamaan rumus:

$$R_i = t_p - t_i \quad (1)$$

dengan:

R_i adalah nilai *recency* pada pelanggan ke- i

t_i adalah waktu transaksi yang terakhir kali dilakukan oleh pelanggan ke- i

t_p adalah waktu acuan yang ditetapkan untuk penelitian

b. *Frequency* adalah total jumlah transaksi yang dilakukan selama periode tertentu. Variabel *frequency* diperoleh dari jumlah transaksi yang dilakukan oleh masing-masing pelanggan dengan mengagregasikan semua data berdasarkan setiap pelanggan. Nilai *frequency* didapatkan dengan menggunakan persamaan rumus:

$$F_i = \sum_{j=1}^n a_{i,j} \quad (2)$$

dengan:

F_i adalah nilai *frequency* pada pelanggan ke- i

i adalah banyaknya pelanggan

n adalah banyaknya transaksi selama periode penelitian

$a_{i,j}$ adalah pemesanan yang dilakukan oleh pelanggan ke- i

- c. *Monetary* adalah total nilai produk dalam bentuk uang pada periode tertentu. Variabel *monetary* diperoleh dari jumlah total pendapatan yang diperoleh perusahaan berdasarkan masing-masing pelanggan. Nilai *monetary* didapatkan dengan persamaan rumus:

$$M_i = \sum_{j=1}^n b_{i,j} \quad (3)$$

dengan:

M_i adalah nilai *monetary* pada pelanggan ke- i

i adalah banyaknya customer

n adalah banyaknya transaksi selama periode penelitian

$b_{i,j}$ adalah pengeluaran yang dilakukan pelanggan ke- i dalam bertransaksi selama periode penelitian

2.3. Data Mining

Data mining merupakan proses pencarian ataupun penambangan informasi baru dengan mencari pola-pola yang menarik dan tersembunyi (*hidden pattern*) dari suatu kumpulan data yang berukuran besar yang tersimpan dalam suatu basis data, data warehouse, atau tempat penyimpanan data lainnya (Tan *et al.*, 2006). Tahap-tahap *data mining* adalah sebagai berikut :

a. Data Preprocessing

Proses ini bertujuan untuk mentransformasikan data input ke dalam format yang sesuai untuk kemudian dianalisa. Tahap-tahap *preprocessing* yang dilakukan antara lain:

1. *Data cleaning*, adalah proses analisa kualitas dari suatu data dengan cara mengubah, mengoreksi, atau menghapus data-data yang salah, tidak lengkap, tidak akurat, atau memiliki format yang salah.
2. *Data Selection*, merupakan proses pemilihan dan meminimalkan jumlah data yang digunakan untuk proses *data mining* dengan tetap merepresentasikan data aslinya.
3. Normalisasi *Min-Max*

Normalisasi *Min-Max* merupakan salah satu metode dalam *data transformation* dengan melakukan perubahan linier terhadap data asli yaitu mengubah skala data kedalam rentang yang lebih kecil yaitu [0,1] (Li *et al.*, 2011). Normalisasi *Min-Max* dihitung dengan menggunakan persamaan berikut (Suyanto, 2017) :

$$x'_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (4)$$

dengan:

x'_i adalah nilai baru hasil normalisasi

x_i adalah nilai awal sebelum dinormalisasi

x_{min} adalah nilai terkecil atribut

x_{max} adalah nilai terbesar atribut

b. Data Mining

Proses ini bertujuan untuk mendapatkan pola-pola dan informasi yang tersembunyi di dalam basis data. Ada beberapa teknik yang dapat digunakan dalam *data mining* untuk mendapatkan pola-pola dan informasi tersembunyi, yaitu *classification*, *neural network*, *decision tree*, *genetic algorithm*, *clustering*, OLAP (*Online Analytical Processing*), dan *association rules*.

c. Post Processing

Proses ini bertujuan untuk memastikan hanya hasil yang valid dan berguna yang dapat digunakan oleh pihak yang berkepentingan. Contoh dari proses ini adalah proses evaluasi dan visualisasi, yaitu proses untuk menganalisa dan mengeksplorasi data dan hasil dari proses data mining dari berbagai sudut pandang.

1. *Pattern evaluation*, merupakan proses untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan.
2. *Knowledge presentation*, adalah proses visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

2.4. Analisis Kluster

Analisis kluster adalah istilah yang diberikan pada sebuah teknik yang digunakan untuk mengidentifikasi objek atau individu yang serupa dengan memperhatikan beberapa kriteria. Fokusnya adalah untuk menentukan cara objek atau individu seharusnya digolongkan untuk memastikan adanya kemiripan anggota dalam satu kelompok dan adanya perbedaan antar kelompok. Sebuah kluster harus memiliki homogenitas internal (dalam satu kluster) yang tinggi dan heterogenitas eksternal (antar kluster) yang tinggi (Kuncoro, 2009).

2.5. Asumsi Analisis Kluster

Terdapat dua asumsi dalam analisis kluster yaitu sampel yang mewakili populasi dan multikolinearitas (Fathia *et al.*, 2016).

a. Sampel Mewakili Populasi

Uji Kaiser-Mayer-Olkin (KMO) digunakan untuk melihat syarat kecukupan suatu sampel. Jika nilai KMO berkisar antara 0,5 sampai 1 maka sampel dapat digunakan mewakili populasi atau sampel representatif. Rumus dari KMO adalah sebagai berikut (Ningrat *et al.*, 2016):

$$KMO = \frac{\sum_{j=1}^W \sum_{k=1, k \neq j}^W r^2_{x_j x_k}}{\sum_{j=1}^W \sum_{k=1, k \neq j}^W r^2_{x_j x_k} + \sum_{j=1}^W \sum_{k=1, k \neq j}^W \rho^2_{x_j x_k, x_1}} \quad (5)$$

dengan:

KMO adalah nilai Uji Kaiser-Mayer-Olkin

W

adalah banyak variabel

$r^2_{x_j x_k}$ adalah korelasi antara variabel x_j dan x_k

$\rho^2_{x_j x_k, x_1}$ adalah korelasi parsial antara variabel x_j dan x_k

b. Multikolinearitas

Multikolinieritas adalah adanya hubungan linier yang sempurna atau pasti di antara beberapa atau semua variabel (Gujarati, 2009). Sebaiknya multikolinieritas ini tidak terjadi di antara variabel-variabel. Salah satu cara identifikasi adanya multikolinieritas adalah dengan menghitung nilai Varians Inflation Factor (VIF) dengan rumus sebagai berikut (Gujarati, 2009):

$$VIF = \frac{1}{1 - R^2} \quad (6)$$

Dengan R^2 adalah nilai koefisien determinasi variabel dependen dengan variabel independen. Multikolinearitas terjadi apabila nilai $VIF > 10$.

2.6. Konsep Jarak

Analisis kluster membuat pengelompokan objek berdasarkan jarak antara pasangan objek. Jarak merupakan ukuran yang digunakan untuk mengukur kemiripan dari suatu objek. Pada proses pemilihan ini, dapat dipilih salah satu dari beberapa jarak yang biasa digunakan salah satunya adalah jarak Euclidean. Jarak Euclidean antara kluster objek ke- i dan kluster objek ke- j dari p variabel didefinisikan sebagai berikut (Santosa, 2007):

$$d(x_i, x_j) = \left[\sum_{k=1}^c (x_{ik} - x_{jk})^2 \right]^{\frac{1}{2}} \quad (7)$$

dengan

$d(x_i, x_j)$ adalah jarak euclidean objek data ke-i dan objek data ke-j

p adalah banyaknya peubah

x_{ij} adalah objek data ke-i pada peubah ke-j

x_{kj} adalah objek data ke-k pada peubah ke-j

2.7. Teori Himpunan Fuzzy

Teori himpunan fuzzy merupakan kerangka matematis yang digunakan untuk menggambarkan ketidakpastian, ketidakjelasan dan ketidaktepatan (Zadeh, 1965). Konsep himpunan fuzzy sejalan dengan himpunan tegas (*hard*), akan tetapi tingkat keanggotaan dari himpunan fuzzy bersifat kontinu yang nilainya dalam interval $[0,1]$. Contohnya diketahui sebuah himpunan $X = \{x_i\}$ dengan $i = 1, 2, \dots, n$. Suatu himpunan fuzzy dalam X didefinisikan sebagai tingkat keanggotaan (u_{ik}) yang berasosiasi dengan setiap x_i pada interval $[0,1]$ dengan $k = 1, 2, \dots, c$ adalah banyaknya himpunan. Nilai (u_{ik}) menunjukkan tingkat keanggotaan x_i , semakin besar nilai keanggotaan atau mendekati nilai satu maka semakin tinggi tingkat keanggotaan x_i .

2.8. Fuzzy Clustering

Menurut Bezdek dalam (Sanusi *et al.*, 2018), Fuzzy Clustering merupakan salah satu metode analisis kluster dengan mempertimbangkan tingkat keanggotaan yang mencakup himpunan fuzzy sebagai dasar pembobot bagi pengelompokan. Metode ini merupakan pengembangan dari metode *partitioning* data dengan pembobotan fuzzy. Keunggulan utama fuzzy clustering adalah dapat memberikan hasil pengelompokan bagi objek-objek yang tersebar tidak teratur, karena jika terdapat suatu data yang penyebarannya tidak teratur maka terdapat kemungkinan suatu titik data mempunyai sifat atau karakteristik dari kluster lain.

2.9. Algoritma Fuzzy C-Means Clustering

Fuzzy C-Means adalah suatu teknik pengelompokan data yang keberadaan tiap-tiap data dalam suatu kelompok ditentukan oleh nilai atau derajat keanggotaan tertentu. Fuzzy C-Means menerapkan pengelompokan fuzzy, dimana setiap data dapat menjadi anggota dari beberapa kluster dengan derajat keanggotaan yang berbeda-beda pada setiap kluster. Dalam teori fuzzy, keanggotaan sebuah data tidak diberikan nilai secara tegas dengan nilai 1 (menjadi anggota) dan nilai 0 (tidak menjadi anggota), melainkan dengan suatu nilai derajat keanggotaannya yang jangkauan nilainya 0 sampai 1. Algoritma dari Fuzzy C-Means kluster adalah sebagai berikut (Haqiqi dan Kurniawan, 2015) :

1. Memasukkan data x_{ij} yang akan dikluster ke dalam sebuah matriks, dimana matriks berukuran $n \times w$, dengan n adalah jumlah data yang akan dikluster dan w adalah banyak atribut atau variabel.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1w} \\ x_{21} & x_{22} & \dots & x_{2w} \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nw} \end{bmatrix}$$

dengan:

x_{ij} adalah objek data ke-i ($i = 1, 2, \dots, n$) pada variabel ke-j ($j = 1, 2, \dots, w$)

n adalah banyaknya objek data

w adalah banyaknya variabel

- Menentukan banyak kelompok (c), fuzzifier (m), maksimum iterasi (Max Iter), perubahan nilai fungsi objektif terkecil yang diharapkan (ϵ), fungsi objektif awal ($J_0 = 0$), dan iterasi awal ($t = 1$).
- Membangkitkan matriks bilangan random u_{ik} dengan i merupakan banyak data dan k merupakan banyak kelompok sebagai elemen-elemen awal matriks keanggotaan awal U .

$$U^{(0)} = \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1c} \\ u_{21} & u_{22} & \dots & u_{2c} \\ \vdots & \vdots & \dots & \vdots \\ u_{n1} & u_{n2} & \dots & u_{nc} \end{bmatrix} \quad (8)$$

u_{ik} adalah nilai keanggotaan objek data ke- i ($i = 1, 2, \dots, n$) dengan pusat kelompok ke- k ($k = 1, 2, \dots, c$)

n banyaknya objek data

c banyaknya pusat kelompok

- Menghitung pusat kelompok ke- i dengan persamaan :

$$v_k = \frac{\sum_{i=1}^n \sum_{j=1}^w (u_{ik})^m x_{ij}}{\sum_{i=1}^n (u_{ik})^m} \quad (9)$$

dengan :

v_k adalah nilai pusat kelompok ke- k

u_{ik} adalah nilai keanggotaan objek ke- i dengan pusat kelompok ke- k

x_{ij} adalah objek data ke- i pada variabel ke- j

n adalah banyaknya objek data

m adalah fuzzifier

w adalah banyak variabel

- Menghitung fungsi objektif pada iterasi ke- t dengan persamaan :

$$J_t = \sum_{i=1}^n \sum_{k=1}^c \left((u_{ik})^m \left[\sum_{j=1}^w d_{ik}^2(x_{ij}, v_{kj}) \right] \right) \quad (10)$$

dengan :

J_t adalah fungsi objektif pada iterasi ke- t

x_{ij} adalah data ke- i pada variabel ke- j

v_{kj} adalah nilai pusat kelompok ke- k pada variabel ke- j

u_{ik} adalah nilai keanggotaan objek data ke- i dengan pusat kelompok ke- k

$d_{ik}^2(x_{ij}, v_{kj})$ adalah jarak antara objek data ke- i dengan pusat kelompok ke- k

n adalah banyaknya objek data

c adalah banyaknya pusat kelompok

w adalah banyaknya variabel

m adalah pangkat pembobot (fuzzifier)

- Menghitung perubahan matriks keanggotaan dengan persamaan:

$$u_{ik}^{(t)} = \frac{1}{\sum_{l=1}^c \left(\frac{d_{ik}^2}{d_{il}^2} \right)^{\frac{1}{m-1}}} \quad (11)$$

Nilai d_{ik}^2 dirumuskan dengan :

$$d_{ik}^2 = \|x_{ij} - v_{kj}\|^2 = (x_{ij} - v_{kj})A(x_{ij} - v_{kj})^T \quad (12)$$

Salah satu konsep jarak adalah jarak Euclidean dengan matriks pembobot yang berkorespondensi $A = I$, dengan I adalah matriks identitas berukuran $w \times w$.

Dengan :

u_{ik} adalah nilai keanggotaan objek ke- k dengan pusat kelompok ke- i

d_{ik}^2 adalah jarak antara objek ke- k dengan pusat kelompok ke- i

d_{il}^2 adalah jarak antara objek ke- k dengan pusat kelompok ke- l

m adalah *fuzzifier*

7. Cek kondisi berhenti

a. Jika $|J_t - J_{t-1}| < \varepsilon$ atau $t > MaxIter$ maka berhenti;

b. Jika tidak : $t + 1$, ulangi langkah ke-4

2.10. Pengukuran Kualitas Kluster

Untuk menentukan berapa banyak kluster optimal dalam pengelompokan data digunakan indeks validitas. Xie Benie Index ditemukan oleh Xie dan Benie pada tahun 1991, indeks ini bertujuan untuk mengukur rasio dari total variasi dalam kelompok dan pemisah kelompok. Banyak kelompok optimal diperoleh dengan meminimalkan nilai indeks (Rahmatika *et al.*,2015). Xie Beni indeks dinyatakan dalam persamaan

$$XB = \frac{\sum_{i=1}^n \sum_{k=1}^c \sum_{j=1}^w (u_{ik})^m d_{ik}^2(x_{ij}, v_{kj})}{n \min d_k^2(v_k, v_k)} \quad (13)$$

$d_{ik}^2(x_{ij}, v_{kj})$ merupakan jarak data terhadap pusat kluster, $d_{ik}^2(v_k, v_k)$ merupakan jarak dari pusat kluster v_k ke pusat kluster v_k .

3. METODE PENELITIAN

3.1. Objek Penelitian

Penelitian ini dilakukan di *e-commerce* Tumbas.in pada tahun 2020. Tumbas.in merupakan sebuah *e-commerce* yang memfasilitasi masyarakat untuk berbelanja di pasar tradisional secara *online*. Data yang digunakan pada penelitian ini adalah data sekunder berupa data histori transaksi pelanggan *e-commerce* Tumbas.in pada bulan Maret 2018 sampai dengan Agustus 2018.

3.2. Variabel Penelitian

Data yang digunakan dalam penulisan tugas akhir ini adalah data histori transaksi pelanggan *e-commerce* Tumbas.in. Variabel data yang digunakan pada penelitian ini terdiri dari: Customer ID, Order ID, Product Price Item, Amount, dan Created At, berikut merupakan deskripsi variabel data yang digunakan :

a. Customer ID

Variabel ini menjelaskan mengenai kode unik yang dimiliki setiap pelanggan yang melakukan transaksi belanja di Tumbas.in. Setiap akun atau pelanggan memiliki kode unik masing-masing sebagai penanda pelanggan.

b. Order ID

Variabel ini merupakan variabel yang menjelaskan mengenai kode unik dari setiap pemesanan atau disebut order. Kode unik ini menandai dari setiap pemesanan barang yang dilakukan oleh pelanggan.

c. Product Price Item

Product price item merupakan variabel yang menerangkan tentang harga dari setiap barang yang dijual baik per satuan, bungkus, maupun, kilogram.

d. Amount

Variabel amount merupakan data kuantitatif yang menjelaskan tentang banyaknya kuantitas barang yang dibeli oleh pelanggan pada transaksi yang dilakukan.

e. Created At

Created At berisi informasi mengenai tanggal dan waktu transaksi pembelian yang dilakukan oleh pelanggan.

3.3. Tahapan Penelitian

Penelitian ini bertujuan untuk menerapkan Fuzzy C-Means Kluster dan metode RFM dalam membentuk segmentasi pelanggan pada *e-commerce* Tumbas.in. Adapun tahapan analisis yang dilakukan adalah sebagai berikut:

1. Wawancara pihak *co.founder* Tumbas.in untuk mengetahui secara jelas terhadap permasalahan yang diambil, serta pengambilan data pada database *e-commerce* Tumbas.in
2. Menyiapkan data histori transaksi pelanggan
3. Melakukan *preprocessing data*:
 - a. Data selection : Pemilihan data untuk menentukan atribut ataupun variabel yang akan digunakan sebagai variabel kluster
 - b. Analisis data RFM : Analisis terhadap data transaksi pelanggan untuk menghasilkan variabel yang dapat digunakan pada analisis kluster.
 - d. Normalisasi data : Normalisasi data dilakukan untuk menyetarakan skala nilai variabel *monetary* yang cukup jauh dengan nilai *recency*, dan *frequency*. Metode yang digunakan pada penelitian data ini yaitu normalisasi Min-Max.
4. Melakukan Uji Asumsi
 - a. Uji KMO
 - b. Uji Multikolinearitas
5. Melakukan analisis segmentasi pelanggan tumbas.in dari data yang telah melalui tahap *preprocessing data*. Analisis segmentasi pelanggan tumbas.in menggunakan algoritma Fuzzy C-Means dengan langkah-langkah sebagai berikut:
 - a. Menentukan banyak kelompok (c), fuzzifier (m), maksimum iterasi, perubahan nilai fungsi objektif terkecil yang diharapkan (ϵ), fungsi objektif awal ($J_0 = 0$), dan iterasi awal ($t = 1$).
 - b. Membangkitkan bilangan random u_{ik} dengan i merupakan banyak data dan k merupakan banyak kelompok sebagai elemen awal matriks keanggotaan awal U .
 - c. Menghitung pusat kelompok ke- i dengan Persamaan (9).
 - d. Menghitung fungsi objektif pada iterasi ke- t dengan Persamaan (10).
 - e. Menghitung perubahan matriks keanggotaan dengan Persamaan (11).
 - f. Cek kondisi berhenti:
Jika $|J_t - J_{t-1}| < \epsilon$ atau $t > MaxIter$ maka berhenti;
Jika tidak : $t + 1$, ulangi langkah ke-3
6. Melakukan pengukuran kualitas kluster menggunakan *Xie and Benie index* terhadap kluster yang telah dibentuk dari metode Fuzzy C-Means.
7. Membuat profiling dan interpretasi terhadap kluster-kluster yang telah terbentuk.

4. HASIL DAN PEMBAHASAN

4.1. Preprocessing Data

Tahap *preprocessing data* digunakan untuk memproses data yang masih bersifat tidak terstruktur untuk diolah menjadi bentuk yang lebih sesuai. Proses yang dilakukan berupa *data selection* dan penyesuaian format data. Dalam penelitian tugas akhir ini data yang digunakan adalah data transaksi pelanggan pada *e-commerce* Tumbas.in. Dari data mentah tersebut selanjutnya akan dilakukan proses *data selection* yaitu untuk memilih atribut data yang sesuai dengan analisis RFM. Variabel yang dipilih adalah *customer id*, *order id*, *amount*, *created at*, dan *product price item*. Langkah berikutnya adalah mengubah format tipe data menjadi bentuk format data yang sesuai seperti diberikan pada Tabel 1.

Tabel 1 Hasil Proses Preprocessing Data

Variabel	Sebelum Penyesuaian Format Data		Sesudah Penyesuaian Format Data	
	Format data	Data	Format data	Data
Order Id	<chr>	“001829b0-240a-11e8-85e8,0018e0c0-240a-11e8-917c”, “002b9960-5fae-11e8-a3ec”	<fct>	001829b0-240a-11e8-85e8,0018e0c0-240a-11e8-917c, 002b9960-5fae-11e8-a3ec
Product Price Item	<dbl>	4000, 5000, 3600, 6500, 16000,.....	<dbl>	4000, 5000, 3600, 6500, 16000,.....
Amount	<dbl>	1, 1, 1, 2, 2, 2, 2, 1, 1,.....	<dbl>	1, 1, 1, 2, 2, 2, 2, 1, 1,.....
Customer Id	<chr>	“0017e710-240a-11e8-95cf,002b5c70-5fae11e8-9500”, “00769910-6106-11e8-8360”,.....	<fct>	0017e710-240a-11e8-95cf,002b5c70-5fae11e8-9500, 00769910-6106-11e8-8360,.....
Created At	<dtm>	2018-03-01 05:23:13, 2018-03-01 05:23:13, 2018-...	<date>	2018-03-10, 2018-05-25, 2018-05-26,.....

4.2. Perhitungan Nilai RFM

Setelah data tersebut diseleksi dan dibentuk kedalam format data seperti pada Tabel 1, selanjutnya adalah melakukan analisis RFM. Analisis ini dilakukan untuk dapat membentuk data kedalam variabel penelitian yang dapat merepresentasikan karakteristik pelanggan yaitu *recency frequency dan monetary*.

Tabel 2. Hasil Perhitungan Nilai RFM

No.	Customer Id	Recency	Frequency	Monetary
1.	0017e710-240a-11e8-95cf-cbe97d9da72e	175	2	49.000
2.	002b5c70-5fae-11e8-9500-6199347df8ad	99	5	155.000
3.	00769910-6106-11e8-8360-f91b858e937b	98	26	237.100
4.	0110aa50-6625-11e8-ae36-e53bce414ccb	91	8	77900
5.	011732e0-8629-11e8-965a-3bfc263141ca	50	4	54100

Dari Tabel 2 terlihat bahwa data rekam transaksi pelanggan telah diubah menjadi variabel data *recency*, *frequency*, dan *monetary*. Berdasarkan analisis RFM didapatkan hasil 947 pelanggan dengan nilai variabel *recency frequency dan monetary*. Variabel data tersebut ditahap selanjutnya akan dilakukan normalisasi data sebelum akhirnya menjadi variabel data yang akan dilakukan analisis kluster.

4.3. Normalisasi Min-Max

Nilai-nilai data yang berbeda-beda rentangnya seringkali perlu dinormalisasi atau distandarisasi agar proses data mining tidak bias. Normalisasi dilakukan untuk menyetarakan skala variabel *monetary* yang terlalu jauh dengan variabel *recency* dan *frequency*. Pada penelitian ini metode normalisasi yang digunakan adalah normalisasi *min-max*, dimana suatu x_i pada ketiga variabel yaitu *recency frequency* dan *monetary* dinormalisasi menjadi nilai baru yang berada dalam rentang [0,1].

Tabel 3. Hasil Normalisasi Min-Max

No.	Customer Id	Recency	Frequency	Monetary
1.	0017e710-240a-11e8-95cf-cbe97d9da72e	0,950819672	0,2941176	0,077246283
2.	002b5c70-5fae-11e8-9500-6199347df8ad	0,535519126	0,11764706	0,249353588
3.	00769910-6106-11e8-8360-f91b858e937b	0,530054645	0,73529412	0,381221719
4.	0110aa50-6625-11e8-ae36-e53bce414ccb	0,491803279	0,20588235	0,123949580
5.	011732e0-8629-11e8-965a-3bfc263141ca	0,267759563	0,08823529	0,085488041

4.4. Uji Asumsi

Terdapat dua asumsi dalam analisis kluster yaitu uji kecukupan sampel dan uji multikolinearitas, sebagai berikut:

1. Uji Kecukupan Sampel

Dari proses komputasi didapatkan nilai KMO sebesar $\geq 0,5$ pada keseluruhan variabel *recency frequency* dan *monetary*, oleh karna itu maka data tersebut dapat dikatakan mewakili populasi atau representatif.

2. Uji Multikolinearitas

Dari proses yang dijalankan didapatkan nilai VIF sebesar < 10 pada keseluruhan variabel *recency frequency* dan *monetary*, oleh karna itu maka setiap variabel pada data tersebut tidak terjadi multikolinieritas.

4.5. Algoritma Fuzzy C-Means Clustering

Proses *clustering* dengan algoritma Fuzzy C-Means dilakukan dengan bantuan *software* Ri386.3.6.2. Data yang digunakan adalah data dari hasil analisis RFM yang telah dinormalisasi. Pangkat fuzzyfier yang digunakan adalah $m=2$, untuk batas *error* terkecil (ϵ) = 10^{-3} dan maksimum iterasi sebanyak 1000. Proses analisis kluster dilakukan sebanyak 6 kali dimulai dengan jumlah kluster sebanyak 2 sampai dengan 7 kluster. Jumlah kluster optimum dari proses *clustering* ditentukan menggunakan indeks validitas Xie-Beni.

Tabel 4. Hasil Perhitungan Jumlah Kluster Optimum

No.	Jumlah Kluster	Nilai Xie-Beni Index
1.	2	0,1938484
2.	3	0,2955140
3.	4	0,1717239
4.	5	0,2037329
5.	6	0,2588580
6.	7	0,3752369

Berdasarkan Tabel 4 dapat dilihat bahwa didapat jumlah kluster yang paling optimum adalah 4 kluster, karena memiliki nilai Xie-Beni Index terkecil yaitu sebesar 0,1717239. Jadi, jumlah kluster yang akan dianalisis pada penelitian ini adalah sebanyak 4 kluster. Hasil analisis kluster dengan jumlah 4 kluster adalah sebagai berikut.

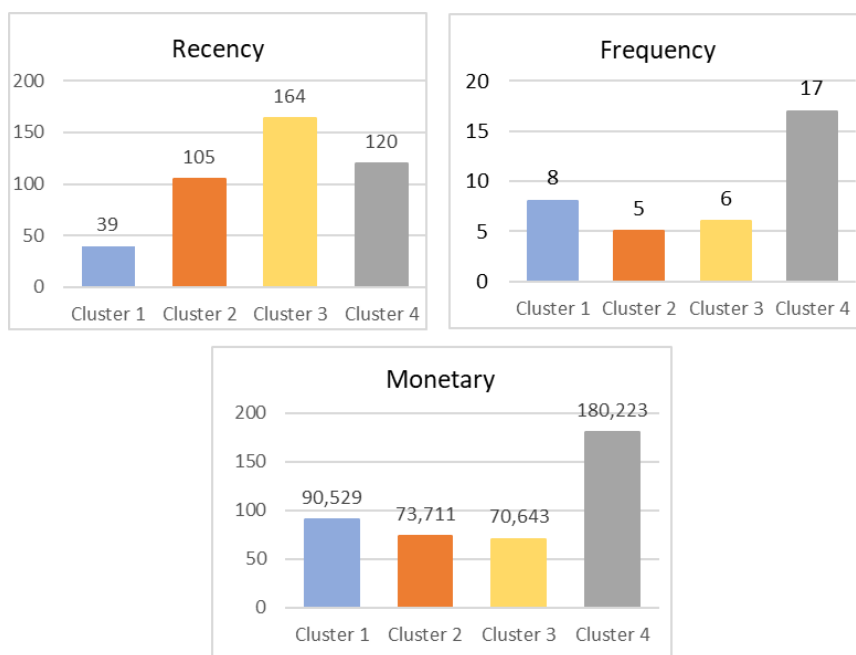
Tabel 5. Hasil Analisis Kluster Dengan Jumlah Kluster Optimum

No.	Customer Id	Kluster	Derajat Keanggotaan			
			Kluster 1	Kluster 2	Kluster 3	Kluster 4
1.	0017e710-240a-11e8-95cf-cbe97d9da72e	4	0,0313	0,1082	0,0605	0,8000
2.	002b5c70-5fae-11e8-9500-6199347df8ad	2	0,0296	0,9140	0,0297	0,0267
3.	00769910-6106-11e8-8360-f91b858e937b	3	0,1535	0,1610	0,5618	0,1238
4.	0110aa50-6625-11e8-ae36-e53bce414ccb	2	0,0931	0,7750	0,0852	0,0467
5.	011732e0-8629-11e8-965a-3bfc263141ca	1	0,7572	0,1563	0,0500	0,0365
⋮	⋮	⋮	⋮	⋮	⋮	⋮
944.	ff51b360-525e-11e8-957d-774da2d709c5	2	0,0277	0,8548	0,0409	0,0766
945.	ffc9e4b0-24ef-11e8-bc5b-51b074f0d087	4	0,0106	0,0391	0,0258	0,9246
946.	ffd96270-6c67-11e8-bb40-4df32134d972	2	0,1619	0,7126	0,0650	0,0605
947.	ffe3fe70-4dfc-11e8-822e-773e8f440c85	2	0,0358	0,7393	0,0773	0,1475

Berdasarkan Tabel 5 didapatkan hasil jumlah anggota dari masing-masing kluster adalah, kluster 1 beranggotakan 242 customer, kluster 2 beranggotakan 292 customer, kluster 3 beranggotakan 144, dan kluster 4 beranggotakan 269 customer.

4.6. Interpretasi Segmen

Strategi pemasaran yang tepat dapat dipilih dengan mempertimbangkan hasil dari nilai karakteristik yang telah terbentuk pada masing-masing segmen. Nilai karakteristik tersebut didapatkan dengan perhitungan rata-rata terhadap variabel *recency frequency* dan *monetary* pada setiap kluster yang telah terbentuk. Berikut merupakan interpretasi dan rekomendasi strategi pemasaran yang tepat pada setiap segmen.



Gambar 1. Grafik Nilai Rata-Rata Karakteristik Segmen

- **Segmen 1**

Segmen 1 memiliki anggota sebanyak 242 pelanggan, dengan rata-rata jumlah pengeluaran yang dikeluarkan oleh pelanggan (*monetary*) sebesar Rp.90.529. Rata-rata jumlah transaksi yang dilakukan oleh pelanggan (*frequency*) adalah 8 kali transaksi perpelanggan. Rata-rata rentang waktu transaksi terakhir kali yang dilakukan oleh pelanggan (*recency*) adalah 39 hari yang lalu. Hal ini menunjukkan bahwa segmen 1 memiliki pelanggan yang mengeluarkan uang relatif cukup besar pada setiap transaksi yang dilakukan, kemudian pelanggan juga cukup sering dalam melakukan transaksi pembelian. Segmen 1 dapat dikategorikan segmen pelanggan yang sedang bertumbuh menjadi pelanggan yang loyal. Oleh karena itu strategi pemasaran yang tepat untuk digunakan adalah dengan melakukan strategi pemasaran yang bertujuan untuk meningkatkan frekuensi pembelian dan jumlah pembelian dalam melakukan transaksi contohnya adalah dengan memberikan penawaran promosi *up selling* ataupun *cross selling*.

- **Segmen 2**

Segmen 2 memiliki anggota sebanyak 292 pelanggan, dengan rata-rata jumlah pengeluaran yang dikeluarkan oleh pelanggan (*monetary*) sebesar Rp.73.711. Rata-rata jumlah transaksi yang dilakukan oleh pelanggan (*frequency*) adalah 5 kali transaksi perpelanggan. Rata-rata rentang waktu transaksi terakhir kali yang dilakukan oleh pelanggan (*recency*) adalah 105 hari yang lalu. Hal ini menunjukkan bahwa segmen 2 memiliki pelanggan yang mengeluarkan uang relatif cukup kecil pada setiap transaksi yang dilakukan jika dibandingkan segmen yang lain, kemudian pelanggan juga tergolong jarang dalam melakukan transaksi pembelian. Segmen 2 dapat dikategorikan segmen pelanggan yang tergolong masih mencoba-coba dalam membeli produk pada e-commerce tumbas.in, hal itu dapat dilihat dari intensitas transaksi dan jumlah besaran uang yang dikeluarkan untuk transaksi masih cenderung relatif rendah. Oleh karena itu strategi pemasaran yang tepat untuk digunakan adalah dengan melakukan strategi pemasaran yang bertujuan untuk menarik minat pelanggan dalam melakukan transaksi pembelian contohnya adalah dengan memberikan penawaran promosi dan diskon kepada para pengguna baru aplikasi tumbas.in.

- **Segmen 3**

Segmen 3 memiliki anggota sebanyak 269 pelanggan, dengan rata-rata jumlah pengeluaran yang dikeluarkan oleh pelanggan (*monetary*) sebesar Rp.70.643. Rata-rata jumlah transaksi yang dilakukan oleh pelanggan (*frequency*) adalah 6 kali transaksi perpelanggan. Rata-rata rentang waktu transaksi terakhir kali yang dilakukan oleh pelanggan (*recency*) adalah 164 hari yang lalu. Hal ini menunjukkan bahwa segmen 3 juga memiliki pelanggan yang mengeluarkan uang relatif cukup kecil pada setiap transaksi yang dilakukan, kemudian pelanggan juga tergolong jarang dalam melakukan transaksi pembelian. Segmen 3 juga dapat dikategorikan segmen pelanggan yang tergolong masih mencoba-coba dalam membeli produk pada e-commerce tumbas.in, hal itu dapat dilihat dari intensitas transaksi dan jumlah besaran uang yang dikeluarkan untuk transaksi masih cenderung relatif rendah. Oleh karena itu strategi pemasaran yang tepat untuk digunakan adalah dengan melakukan strategi pemasaran yang bertujuan untuk menarik minat pelanggan dalam melakukan transaksi pembelian contohnya adalah dengan memberikan penawaran promosi dan diskon kepada para pengguna baru aplikasi tumbas.in.

- **Segmen 4**

Segmen 4 memiliki anggota sebanyak 144 pelanggan, dengan rata-rata jumlah pengeluaran yang dikeluarkan oleh pelanggan (*monetary*) sebesar Rp.180.223. Rata-rata jumlah transaksi yang dilakukan oleh pelanggan (*frequency*) adalah 17 kali transaksi

perpelanggan. Rata-rata rentang waktu transaksi terakhir kali yang dilakukan oleh pelanggan (*recency*) adalah 120 hari yang lalu. Hal ini menunjukkan bahwa segmen 4 memiliki pelanggan yang mengeluarkan uang sangat besar dibandingkan segmen lain pada setiap transaksi yang dilakukan, kemudian pelanggan juga tergolong sering dalam melakukan transaksi pembelian. Segmen 4 dapat dikategorikan segmen pelanggan yang terbaik dan loyal dalam membeli produk pada *e-commerce* Tumbas.in, hal itu dapat dilihat dari intensitas transaksi dan jumlah besaran uang yang dikeluarkan untuk transaksi sangatlah tinggi. Oleh karena itu strategi pemasaran yang tepat untuk digunakan adalah dengan melakukan strategi pemasaran yang bertujuan untuk menjaga dan mempertahankan loyalitas pelanggan contohnya adalah dengan memberikan reward dan loyalty program.

5. KESIMPULAN

Fuzzy C-Means Kluster dan metode RFM memiliki performa dan hasil yang cukup efektif dalam melakukan segmentasi pelanggan dengan menggunakan data transaksi pelanggan. Data transaksi pelanggan yang bersifat tidak terstruktur dapat diubah kedalam bentuk variabel penelitian *recency*, *frequency*, dan *monetary* dengan menggunakan metode RFM. Dari analisis RFM dihasilkan 947 data pelanggan dengan nilai *recency*, *frequency*, dan *monetary* yang dapat merepresentasikan karakteristik pelanggan berdasarkan aktifitas belanjanya pada *e-commerce* Tumbas.in.

Proses klustering menggunakan Fuzzy C-Means Kluster dilakukan pada 947 data pelanggan RFM dengan jumlah kluster sebanyak 2 hingga 7 kluster. Selanjutnya dalam menentukan jumlah kluster optimum digunakan nilai Xie-Beni Index, dan didapatkan jumlah kluster terbaik adalah empat kluster ($k=4$) yang mempunyai nilai index paling kecil yaitu sebesar 0,1717239. Dalam penelitian ini menghasilkan 4 kluster sebagai segmen pelanggan pada *e-commerce* Tumbas.in.

DAFTAR PUSTAKA

- Haqiqi, B., Kurniawan, R. 2015. *Analisis Perbandingan Metode Fuzzy C-Means Dan Subtractive Fuzzy C-Means*. Media Statistika, Hal. 59-67.
- Hughes, A. 1994. *Strategic database marketing masterplan for starting and managing a profitable, customer-based marketing program*. Chicago: Probush Publishing.
- Ningrat, D., Maruddani, D., Wuryandari, T. 2016. *Analisis Cluster Dengan Algoritma K-Means dan Fuzzy C-Means Clustering Untuk Pengelompokan Data Obligasi*. Jurnal Gaussian, Vol.5 No.4 Hal. 641-650.
- Santosa, B. 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Sanusi, W., Zakky, A., Afni, B. 2018. *Analisis Fuzzy C-Means dan Penerapannya Dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Faktor-faktor Penyebab Gizi Buruk*.
- Saputra, D., Riksakomara, E. 2018. *Implementasi Fuzzy C-Means dan Model RFM Untuk Segmentasi Pelanggan*. Jurnal Teknik ITS, Vol.7 No.1 2337-3520.
- Suyanto, D. 2017. *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Penerbit Informatika.
- Tan, P., Steinbech, M., Kumar, V. 2006. *Introduction to data mining*. Boston: Pearsong Education,Ltd.