

SEGMENTASI E-COMMERCE DENGAN CLUSTER K-MEANS DAN FUZZY C-MEANS (Studi Kasus: Media Sosial di Indonesia yang diunduh di Play Store)

Fariska Atha Dewa¹, Maria Titah Jatipaningrum^{2*}

^{1,2} Jurusan Statistika, FST, Institut Sains & Teknologi AKPRIND Yogyakarta

E-mail: Fatha_dewa@yahoo.co.id¹, titahjp@akprind.ac.id²

Abstract: E-commerce segmentation is grouping in business activities involving company and consumer. The process of clustering is very needed in the case of e-commerce segmentation which purposed to know the description of online store segmentation on social media in Indonesia. In this research used cluster K-means and Fuzzy C-means method with research variables the number of followers on Twitter, Instagram, and Facebook on social media. The method of determining the number of clusters using the Elbow cluster method, Silhouette Index, and Gap statistics are three clusters, then the results of clustering with K-means method obtained result, that is 1 of 34 member with average moderate number of followers, cluster 2 of 4 with average little number of followers, and cluster 3 of 2 with the average many number followers, with the Silhouette Index value is 0,8641212 and Partion Coefficient Index amounting to 0,9684215. While the results of clustering with Fuzzy C-means method obtained results, that is cluster 1 for 34 with average little number of follower, cluster 2 for 2 with the average many number of follower, and cluster 3 for 4 with average moderate number of followers, with Silhouette Index value of 0,9527799 and Partion Coefficient Index of 0,9685199. The results of the comparison of the best method is the Fuzzy C-means method because value of SI 0,9527799 and PCI 0,9685199 which is greater than the value of SI and PCI in the K-Means method.

Keywords: K-Means, Fuzzy C-Means, Validity

Abstrak: Segmentasi e-commerce merupakan pengelompokan dalam kegiatan berbisnis yang melibatkan perusahaan dan konsumen. Proses pengelompokan sangat diperlukan pada kasus segmentasi e-commerce yang bertujuan untuk mengetahui gambaran segmentasi toko online pada media sosial di Indonesia. Pada penelitian ini digunakan metode K-means cluster dan Fuzzy c-means cluster dengan variabel penelitian jumlah followers pada media sosial twitter, instagram, dan facebook. Metode penentuan jumlah cluster menggunakan metode Elbow, Silhouette, dan Gap Statistik menghasilkan sebanyak tiga klaster. Pengklasteran dengan metode K-Means diperoleh hasil yaitu klaster 1 sebesar 34 anggota dengan rata-rata jumlah followers sedang, klaster 2 sebesar 4 anggota dengan rata-rata jumlah followers sedikit, dan klaster 3 sebesar 2 anggota dengan rata-rata jumlah followers banyak, dengan nilai Silhouette Index sebesar 0,8641212 dan Partion Coefficient Index sebesar 0,9684215. Sedangkan hasil pengkasteran dengan metode Fuzzy c-means diperoleh hasil yaitu klaster 1 sebesar 34 anggota dengan rata-rata jumlah followers sedikit, klaster 2 sebesar 2 anggota dengan rata-rata jumlah followers banyak, dan klaster 3 sebesar 4 anggota dengan rata-rata jumlah sedang, dengan nilai Silhouette Index sebesar 0,9527799 dan Partion Coefficient Index sebesar 0,9685199. Hasil perbandingan metode terbaik antara metode K-Means dan metode Fuzzy c-means adalah metode Fuzzy c-means karena nilai SI 0,9527799 dan PCI 0,9685199 yang lebih besar dari pada nilai SI dan PCI pada metode K-Means.

Kata kunci: K-Means, Fuzzy C-Means, Validitas

1. PENDAHULUAN

E-commerce adalah suatu kegiatan bisnis dengan memanfaatkan penggunaan jaringan internet. *E-commerce* dapat dilakukan tanpa batas, hal ini disebabkan karena Elektronik Commerce merupakan jaringan komputerisasi yang sifatnya sangat global, yakni dapat diakses di seluruh belahan dunia pada waktu yang tak terbatas atau dengan kata lain *on-line* 24 jam setiap hari tanpa batas. Segala informasi dapat diakses kapanpun, di manapun dan saat apapun, sehingga dengan kecanggihan jaringan computer. *E-commerce* di Indonesia semakin mudah dilakukan melihat banyaknya pengguna internet dan semakin mudahnya proses transaksi yang dilakukan pelanggan kepada pelaku *e-commerce*. Hal ini bisa dilihat dari segi masyarakat yang

*Corresponding author's email: : titahjp@akprind.ac.id

mengikuti, menyukai, dan menginstal suatu aplikasi toko online dari internet. Didalam persaingan antar perusahaan *e-commerce* banyak berbagai perusahaan *e-commerce* ternama di Indonesia berlomba untuk berinovasi demi meningkatkan performa bisnisnya. Banyak perusahaan *e-commerce* melakukan meningkatkan pelayanan-pelayanan yang mereka suguhkan demi memperoleh pasar atau konsumen melebihi pesaing mereka bahkan mereka juga memberikan diskon berupa *free* ongkos kirim atau pun diskon harga barang untuk menarik pelanggan/konsumen untuk bisnis *e-commerce*.

Proses *clustering* adalah proses pengelompokan data ke dalam kelompok berdasarkan parameter tertentu sehingga obyek-obyek dalam sebuah *cluster* memiliki tingkat kemiripan yang tinggi satu sama lain dan sangat tidak mirip dengan obyek yang lain pada *cluster* yang berbeda. Sampai saat ini, para ilmuwan masih terus melakukan berbagai usaha untuk melakukan perbandingan model *cluster* sehingga dapat menghasilkan *cluster* yang cocok digunakan pada sebuah penelitian. *K-Means* adalah suatu metode data mining yang melakukan pengelompokan data dengan sistem partisi. Sedangkan *Fuzzy c-Means* merupakan perkembangan dari metode *k-means* dengan memperhitungkan bahwa data dapat tergabung ke dalam ke dalam beberapa *cluster* dengan tingkat keanggotaan yang berbeda-beda. Terlepas dari permasalahan segmentasi persaingan *e-commerce*, berdasarkan kelompok-kelompok yang menjadi sebuah aspek utama dalam jual beli online perlu dilakukannya sebuah *cluster* data. *K-means* dan *fuzzy c-means* jika dilihat dari beberapa riset terdahulu mampu memberikan hasil *cluster* terbaik. Kedua metode tersebut sama-sama memiliki hasil yang signifikan serta mempunyai beberapa perbedaan dalam hal pola *cluster*, oleh karena itu dilakukan perbandingan pada pemodelan *k-means* dan *fuzzy c-means* untuk melihat hasil *clusterisasi*.

Untuk mengetahui *K-Means Clustering* dan *Fuzzy C-Means Clustering* sebagai penelitian dalam hal Pengelompokan data, diantaranya penelitian yang dilakukan oleh Aditya Ramadhan (2017) dengan judul “Perbandingan *K-Means* dan *Fuzzy C-Means* untuk Pengelompokan Data *User Knowledge Modeling*” dan penelitian yang dilakukan oleh Nur indah selviana (2016) dengan judul “Analisis Perbandingan *K-Means* dan *Fuzzy C-Means* untuk Pemetaan Motivasi Belajar Mahasiswa” Penelitian ini menggunakan algoritma *K-Means* dan *Fuzzy C-Means*. Berdasarkan latar belakang tersebut, peneliti mengambil judul “Segmentasi E-commerce Dengan *K-means* Dan *Fuzzy C-Means*”.

2. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder jumlah *followers* pada toko online media sosial Facebook, Twitter, dan Instagram. Pengambilan data dilakukan pada 1 Juni 2018 dengan metode *sampling non probability* yaitu jenis *purposive sampling*. Data yang digunakan sebanyak 40 toko online yang sudah masuk dalam aplikasi *play store*. Dimulai dengan analisis deskriptif yaitu untuk memberikan gambaran umum tentang data yang telah diperoleh dan dapat menjadi acuan untuk melihat karakteristik data.

Penentuan jumlah cluster optimal

Metode yang digunakan dalam penentuan jumlah cluster optimal adalah metode Elbow, Silhouette dan Statistic Gap.

Metode Elbow

Metode *Elbow* menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik. Hasil persentase yang berbeda dari setiap nilai cluster dapat ditunjukkan dengan menggunakan grafik sebagai sumber informasinya. Jika nilai cluster pertama dengan nilai cluster kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai *cluster* tersebut yang terbaik. Untuk mendapatkan perbandingannya adalah dengan menghitung *Sum Square Error* (SSE) dari masing-masing nilai *cluster*. Karena semakin besar jumlah *cluster* K maka nilai SSE akan semakin kecil. Berikut adalah rumus SSE pada *K-Means*:

$$SSE = \sum_{k=1}^k \sum_{x_i \in C_k} |x_i - \mu_k|^2 \quad (1)$$

Keterangan:

Sum Square Error (SSE), c_k merupakan kelompok data dalam cluster k , μ_k adalah rata-rata cluster k , dan x_i adalah x pada indeks ke- i .

Algoritma Metode *Elbow* dalam menentukan nilai K pada *K-Means* adalah sebagai berikut:

1. Hitung algoritma *K-means* untuk nilai yang berbeda dari k . Misalnya, dengan memvariasikan k dari 1 hingga 5 kluster.
2. Untuk setiap k , hitung total jumlah kluster kuadrat dalam-dalam (wss)
3. Plot kurva dari wss sesuai dengan jumlah cluster k .
4. Lokasi tikungan (lutut) di plot umumnya dianggap sebagai indikator jumlah kluster yang sesuai. Perhatikan bahwa, metode siku terkadang ambigu. Alternatifnya adalah rata-rata

Metode Silhouette

Metode *Silhouette* adalah salah satu metode untuk menentukan jumlah *cluster* yang dengan melakukan pendekatan rata-rata nilai metode *silhouette* untuk menduga kualitas *cluster* yang terbentuk. Semakin tinggi nilai rata-rata maka akan semakin baik nilai *silhouettenya*. *Silhouette* juga bisa dihitung menggunakan jarak antar *cluster* (a) dan jarak antar *cluster* terdekat (b) untuk setiap sampel yaitu sebagai berikut:

$$s(i) = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (2)$$

Keterangan:

a_i merupakan rata-rata jarak i dengan observasi lain dalam satu *cluster* dan b_i adalah rata-rata jarak antara i dengan observasi pada *cluster* terdekat.

Tahapan penentuan jumlah *cluster* optimal menggunakan metode *silhouette* yaitu sebagai berikut:

1. Hitung algoritma pengelompokan (misalnya, *k-means clustering*) untuk nilai yang berbeda dari k . Misalnya, dengan memvariasikan k dari 1 hingga 10 kluster.
2. Untuk setiap k , hitung *silhouette* rata-rata observasi (avg.sil).
3. Kurva alur avg. Menurut jumlah *cluster* k .
4. Lokasi jumlah max dianggap sebagai nomor yang sesuai dari *cluster* 1, 2, dan 3.

Metode Gap statistic

Metode *Gap statistic* menghitung kebaikan ukuran pengelompokan dengan melihat nilai $\log W$ dengan $\log E$ untuk setiap jumlah *cluster*. *Gap statistic* akan menentukan jumlah *cluster* terbentuk dengan melihat pada grafik *gap statistic*. Berikut adalah Statistik gap untuk k tertentu:

$$Gap_n(k) = E_n^* \log(W_k) - \log(W_k) \quad (3)$$

Keterangan :

E_n^* = ukuran sampel n dan $(W_k) = \log$ rata-rata

Berikut adalah tahapan penentuan jumlah *cluster* optimal menggunakan metode *Gap statistic* yaitu sebagai berikut:

1. Menggabungkan data yang diamati, memvariasikan jumlah kluster dari $k = 1, \dots, k_{max}$, dan hitung total yang sesuai dalam variasi intra-cluster W_k
2. Hasilkan kumpulan data referensi B dengan distribusi seragam acak. Cluster masing-masing dari kumpulan data referensi ini dengan berbagai jumlah kelompok $k = 1, \dots, k_{max}$ dan menghitung total yang sesuai dalam variasi intra-cluster W_{kb}
3. Hitung estimasi statistik gap sebagai penyimpangan nilai W_k yang diamati dari W_{kb} dan Hitung juga standar deviasi dari statistik.
4. Pilih jumlah *cluster* sebagai nilai terkecil dari k sehingga statistik gap berada dalam satu standar deviasi dari celah pada $k + 1$

K-Means Cluster

K-Means merupakan algoritma yang membutuhkan parameter *input* sebanyak k dan membagi sekumpulan n objek kedalam k *cluster* sehingga tingkat kemiripan antar anggota dalam satu *cluster* tinggi sedangkan tingkat kemiripan dengan anggota pada *cluster* lain sangat rendah. Kemiripan anggota terhadap *cluster* diukur dengan kedekatan objek terhadap nilai *mean* pada *cluster* atau disebut sebagai *centroid cluster*. Metode *K-Means* merupakan metode *clustering* yang paling sederhana dan umum. Hal ini dikarenakan *K-Means* mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien. Tetapi metode ini mempunyai kelemahan dalam menganalisis persebaran data serta bergantung pada inisialisasi *centroid*. *K-Means* hanya melihat jarak data ke masing-masing *centroid* pada setiap *cluster*. Bedzek (1981). Berikut adalah algoritma dari metode *K-Means*:

- a. Masukkan data yang akan diklaster.
- b. Tentukan k sebagai jumlah cluster yang akan dibentuk.
- c. Tentukan titik pusat cluster (*centroid*).

$$v = \frac{\sum_{i=1}^n x_i}{n}; i = 1, 2, 3, \dots, n \quad (4)$$

Keterangan:

v sebagai *centroid* pada cluster, x_i adalah data ke- i dan n adalah banyaknya data yang menjadi anggota cluster.

- d. Hitung jarak setiap data ke pusat klaster dengan menggunakan *euclidian distance* menurut Hermastuti dan Suryowati (2013), dengan rumus pada persamaan 2.5 sebagai berikut:

$$d_{euc}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5)$$

Keterangan :

Jarak antara x dan y dinotasikan $d(x, y)$; x_i : data x ke- i ; y_i : data y ke- i dan n adalah banyaknya data

- e. Hitung kembali pusat klaster dengan keanggotaan klaster yang baru, jika pusat klaster tidak berubah maka proses klaster telah selesai, jika belum maka ulangi langkah ke d sampai pusat klaster tidak berubah lagi. Berikut adalah rumus untuk menghitung kembali pusat cluster:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (6)$$

Keterangan :

μ_k = titik *centroid* dari cluster ke- K ; N_k = banyaknya data pada cluster ke- K ; x_q = data ke- q pada cluster ke- K

Fuzzy C-Means Cluster

Fuzzy C-Means adalah metode pengklasteran data yang ditentukan oleh derajat keanggotaan. Tujuan dari *Fuzzy C-Means* adalah untuk mendapatkan pusat *cluster* yang nantinya akan digunakan untuk mengetahui data yang masuk ke dalam sebuah *cluster*. Berikut adalah algoritma *Fuzzy C-Means*:

- a. Masukkan data yang akan diklaster, berupa matriks X berukuran $n \times m$
- b. Tentukan :
 1. Jumlah klaster (c)
 2. Pangkat Pembobot (w)
 3. Maksimum Iterasi (*Max Iter*)
 4. Error Terkecil (ϵ)
 5. Fungsi objektif awal ($P_0=0$)
 6. Iterasi awal ($t=1$)
- c. Bangkitkan bilangan acak (μ_{ik}), dengan $i = 1, 2, \dots, n$; $k = 1, 2, \dots, c$ dan c sebagai elemen-elemen matriks partisi awal.
- d. Hitung pusat klaster ke- k : V_{kj} dengan $k = 1, 2, \dots, c$ dan $j = 1, 2, \dots, m$

$$V_{kj} = \frac{\sum_{i=1}^n ((U_{ik})^w * X_{ij})}{\sum_{i=1}^n (U_{ik})^w} \quad (7)$$

Keterangan:

V_{kj} = Pusat *cluster* ke- k dan atribut ke- j

U_{ik} = derajat keanggotaan sampel data ke- i dan cluster ke- k

X_{ik} = data pada sampel ke- i dan atribut ke- j

- e. Hitung fungsi objektif pada iterasi ke- t , untuk melihat jarak data ke- i terhadap pusat *cluster* dengan rumus:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{j=1}^m (X_{ij} - V_{kj})^2] (U_{ik})^w) \quad (8)$$

- f. Hitung perubahan matriks partisi:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}} \quad (9)$$

dengan $i = 1, 2, \dots, n$ dan $k = 1, 2, \dots, c$

- g. Cek kondisi berhenti:

1. Jika: $(|P_t - P_{t-1}| < \epsilon)$ atau $(t > MaxIter)$ maka berhenti.
2. Jika tidak: $t = t + 1$, ulangi langkah ke-d.

Validitas Silhouette Index

Metode *silhouette index* dapat digunakan untuk validitas algoritma dengan melihat nilai SI. Berikut adalah formula untuk perhitungan *Silhouette Index*:

$$SI = \frac{1}{M} \sum_{i=1}^{m_j} SI_i^j \quad (10)$$

Sementara nilai SI global didapatkan dengan menghitung rata-rata nilai SI dari semua *cluster* seperti pada persamaan berikut :

$$SI = \frac{1}{k} \sum_{j=1}^k SI_j \quad (11)$$

Keterangan:

a_i = rata-rata jarak data ke- i terhadap semua data lainnya dalam satu klaster

b_i = hasil rata-rata jarak data ke- i terhadap semua data dari klaster lain

$d(x_i^j, x_r^j)$ = jarak data ke- i dengan data ke- r dalam satu klaster j

m_j = jumlah data dalam klaster ke- j

k = banyaknya klaster

SI_i^j = rumus *Silhouette Index* ; SI_j = SI untuk setiap klaster

Validitas Partion Coefficient Index

Validitas *Partion Coefficient Index* menghitung koefisien partisi sebagai evaluasi nilai keanggotaan data pada setiap *cluster*. Nilai PCI mengavaluasi nilai derajat keanggotaan. Nilai dalam rentang $[0,1]$, nilai yang semakin besar mendekati 1 mempunyai arti bahwa kualitas *cluster* yang didapat semakin baik. Berikut adalah formula untuk menghitung PCI:

$$PCI = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^k u_{ij}^2 \quad (12)$$

Keterangan:

N adalah jumlah data dalam set data ; K merupakan jumlah klaster ; dan u_{ij} adalah nilai keanggotaan data ke- i pada cluster ke- j .

3. HASIL ANALISIS DAN PEMBAHASAN

Analisis Deskriptif

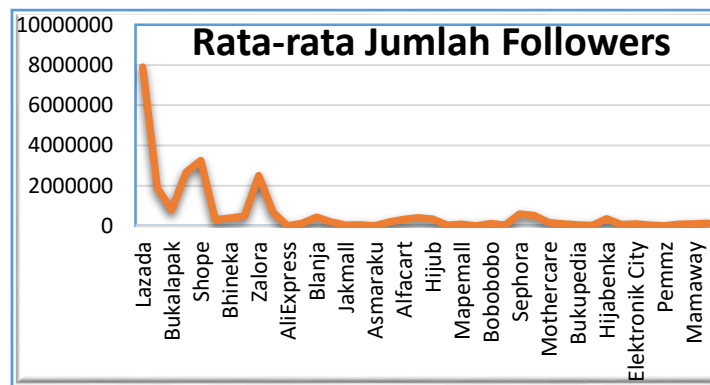
Pada analisis deskriptif dapat dilihat karakteristik jumlah *followers* pada 3 media sosial atau pada masing-masing media social dan pada toko online yang mempunyai jumlah followers 0 adalah toko online yang tidak mempunyai media sosial tetapi sudah masuk kedalam aplikasi *play store*. Hasil analisis pada pembahasan ini ditampilkan dalam bentuk

tabel dan grafik berdasarkan *output* analisis deskriptif (lampiran 2). Berikut adalah tabel analisis deskriptif *e-commerce* Indonesia pada 1 Juni 2018 berdasarkan perhitungan dengan *Microsoft Excel* sebagai berikut:

Tabel 1. Analisis Deskriptif

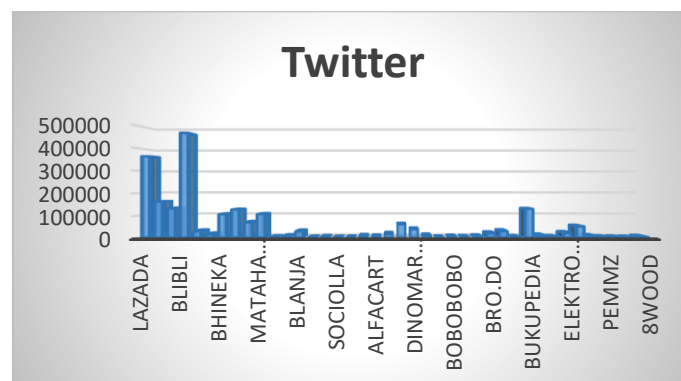
Descriptive Statistics					
Variabel	N	Minimum	Maximum	Mean	Variance
Twitter	40	0	470200	51237.5	9.34
Instagram	40	0	712700	154825	3.57
Facebook	40	0	22768300	2108593	2.24

Tabel 1 menunjukkan gambaran *e-commerce* pada ketiga media sosial yakni mengenai rata-rata jumlah *followers*, nilai maksimum dan minimum serta variansi data *e-commerce*.



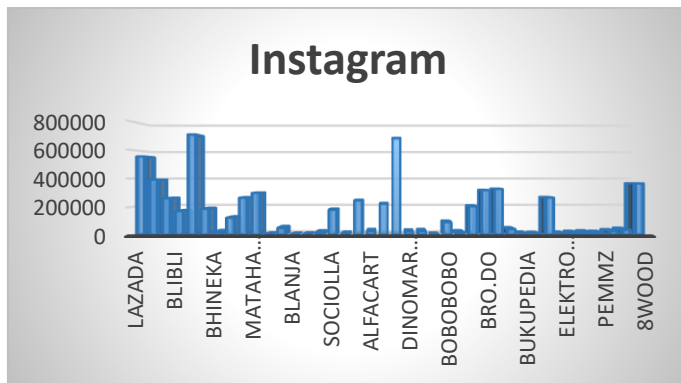
Gambar 1. Rata-rata Jumlah Followers

Gambar 1 menunjukkan rata-rata jumlah *followers* dari ketiga media sosial yang paling tinggi adalah Lazada sebesar 7.896.533 *followers*, sedangkan rata-rata jumlah *followers* dari ketiga media sosial paling rendah adalah Weshop yaitu sebesar 100 *followers*. Kemudian jumlah *followers* pada media sosial Twitter yaitu:



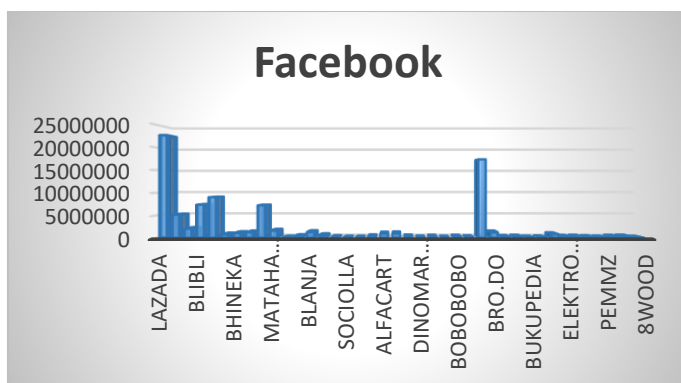
Gambar 2. Jumlah Followers Twitter

Gambar 2 menunjukkan bahwa jumlah *followers* pada media sosial Twitter yang paling tinggi adalah Blibli yaitu sebesar 470.200 *followers* sedangkan yang paling rendah adalah Qoo10 yaitu sebesar 0 *followers*. Kemudian jumlah *followers* pada media sosial Instagram yaitu:



Gambar 3. Jumlah Followers Instagram

Gambar 3 menunjukkan bahwa jumlah *followers* pada media sosial Instagram yang paling tinggi adalah Shopee sebesar 712.700 *followers*, sedangkan yang paling rendah adalah Blanja dan Weshop yaitu sebesar 0 *followers*. Kemudian jumlah *followers* pada media sosial Facebook yaitu:



Gambar 4. Jumlah Followers Facebook

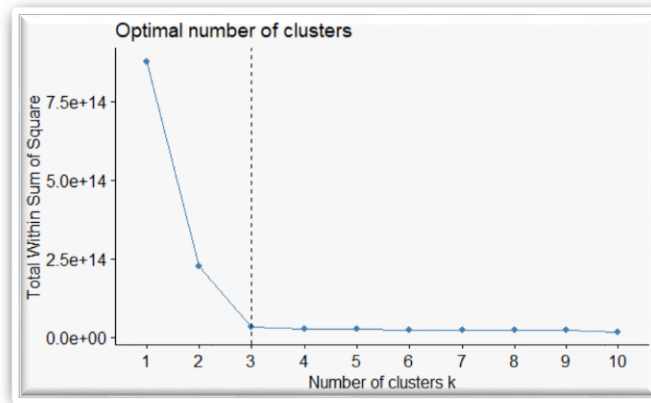
Gambar 4 menunjukkan bahwa jumlah *followers* pada media sosial Instagram yang paling tinggi adalah Lazada sebesar 22.768.300 *followers*, sedangkan yang paling rendah adalah Weshop yaitu sebesar 0 *followers*

Penentuan Jumlah Cluster Optimal

Penentuan nilai *k* optimal bertujuan untuk menentukan banyak *k* kelompok yang paling optimal. Pendekatan metode yang digunakan adalah metode *Elbow*, *Silhouette* dan *Gap_Statistik*.

Metode Elbow

Metode *Elbow* adalah metode dengan kriteria penilaian berdasarkan tingkat kesikuan pada grafik. Berikut adalah hasil dari penentuan jumlah *cluster* optimal berdasarkan grafik metode *Elbow* sebagai berikut:

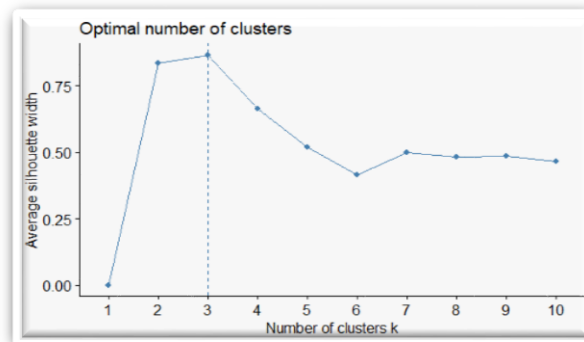


Gambar 5. Grafik Metode Elbow

Gambar 5 menunjukkan garis mengalami patahan yang membentuk *elbow* atau siku pada saat $k = 3$. Maka dengan menggunakan metode ini diperoleh k optimal pada saat berada di $k = 3$ artinya penentuan jumlah *cluster* terbaik adalah 3 *cluster*.

Metode Silhouette

Metode *Silhouette* menentukan jumlah *cluster optimal* dengan melakukan pendekatan rata-rata nilai *silhouette*. Hasil dari penentuan jumlah *cluster optimal* berdasarkan grafik metode *Silhouette* adalah sebagai berikut:

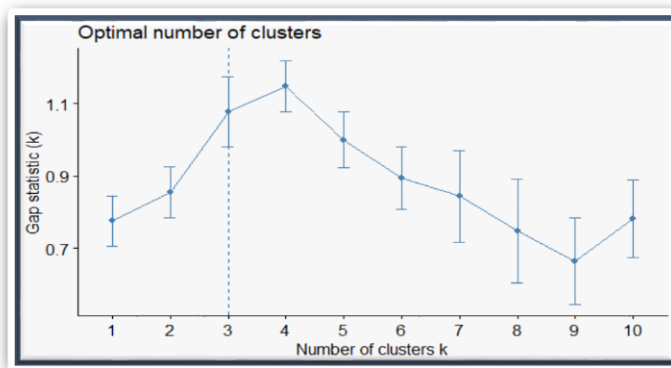


Gambar 6. Grafik Metode Silhouette

Gambar 6 diperoleh banyak *cluster optimal* yang terbentuk pada $k = 3$ yang ditunjukkan dengan nilai rata-rata tertinggi pada $k = 3$, semakin tinggi nilai rata-ratanya maka kualitas dari *cluster* yang terbentuk akan semakin baik.

Metode Gap Statistik

Metode *Gap statistik* menentukan jumlah *cluster optimal* dengan melihat garis titik pada grafik *Gap statistik*. Hasil dari penentuan jumlah *cluster optimal* berdasarkan grafik metode *Gap statistik* adalah sebagai berikut:



Gambar 7. Grafik Metode Gap statistik

Gambar 7 menunjukkan penarikan garis ada di $k = 3$ yang berarti penentuan jumlah cluster optimal adalah 3 cluster.

Clustering Menggunakan Metode K-Means

1. Jumlah Cluster Terbentuk

Tabel 2. Jumlah Cluster Terbentuk

Cluster	1	2	3
Jumlah Anggota	34	4	2

Tabel 2 menunjukkan cluster yang terbentuk adalah sebanyak $k = 3$ dengan jumlah anggota masing-masing cluster yaitu 34, 4, 2 yang berarti hasil *K-Means* dengan $k = 3$ terbentuk cluster 1 sebanyak 34 toko online, cluster 2 sebanyak 4 toko online dan cluster 3 sebanyak 2 toko online.

2. Pusat Cluster Awal

Tabel 3. Pusat Cluster Awal

Cluster Means	Twitter	Instagram	Facebook
1	27911.76	114717.6	456711.8
2	182375.00	383675.0	7162950.0
3	185500.00	378950.0	20083650.0

Tabel 3 menunjukkan nilai rata-rata setiap kelompok yang terbentuk menunjukkan bahwa rata-rata cluster 1 adalah sebagai titik pusat cluster 1, rata-rata cluster 2 adalah sebagai titik pusat cluster 2 dan rata-rata cluster 3 adalah sebagai titik pusat cluster 3.

3. Agregat Cluster Awal

Tabel 4. Agregat Cluster Awal

Group.1	Twitter	Instagram	Facebook
1	27911.76	114717.6	456711.8
2	182375.00	383675.0	7162050.0
3	185500.00	378950.0	20083650.0

Tabel 4 menunjukkan nilai agregat dari masing-masing cluster terhadap variabel yaitu Twitter (27911.76 182375.00 185500.00), Instagram (114717.6 383675.0 378950.0), dan Facebook (456711.0 7162050.0 20083650.0).

4. Hitung Kembali Pusat Cluster

Tabel 5. Pusat Cluster Baru

Cluster Means	Twitter	Instagram	Facebook
1	27911.76	114717.6	456711.8
2	182375.00	383675.0	7162950.0
3	185500.00	378950.0	20083650.0

Tabel 5 menunjukkan nilai pusat *cluster* baru setiap kelompok yang menunjukkan bahwa pusat *cluster* 1 sebagai titik pusat *cluster* 1, pusat *cluster* 2 sebagai titik pusat *cluster* 2 dan pusat *cluster* 3 sebagai titik pusat *cluster* 3. Maka dari perhitungan pusat *cluster* awal dengan pusat *cluster* baru diketahui bahwa pusat *cluster* tidak ada yang berubah yang berarti proses *clustering* telah selesai.

5. Hasil Pengelompokan Anggota

Tabel 6. Hasil Clustering K-Means

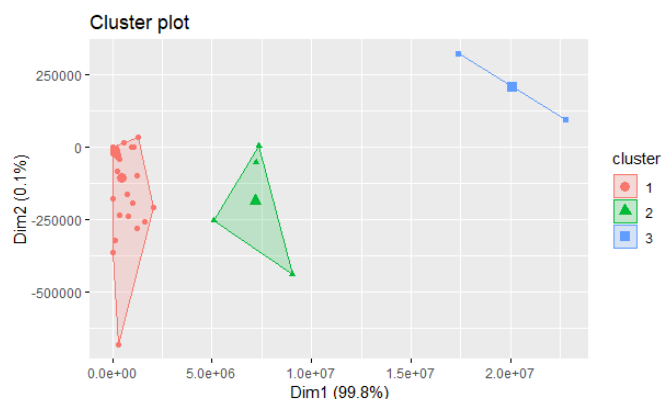
K-means	
Cluster 1	Bukalapak, JD.ID, Bhineka, Eleven, Matahari, Aliexpress, Orami, Blanja, Qoo10, Jakmall, Sociolla, Asmara, Otten_Coffe, Alfacart, Berrybenka, Hijub, Dinomarket, Mapemall, Weshop, Bobobobo, Vip_plaza, Bro.do, Mothercare, Orori, Bukupedia, Tees, Hijabenka, Elektronik_solusion, Elektronik_city, Maskoolin, Pemmz, Muslimarket, Mamaway, dan 8Wood.
Cluster 2	Tokopedia, Blibli, Shope, dan Zalora.
Cluster 3	Lazada dan Shepora.

Cluster 1 beranggotakan 34 anggota dengan karakteristik jumlah followers sedang.

Cluster 2 beranggotakan 4 anggota dengan karakteristik jumlah followers sedikit.

Cluster 3 beranggotakan 2 anggota dengan karakteristik jumlah followers banyak.

6. Plot K-Means



Gambar 8. Plot K-Means

Gambar 8 menunjukkan pola persebaran setiap anggota yang terbentuk dalam tiga *cluster*. Anggota *cluster* 1 adalah yang paling banyak diantara *cluster* 2 dan 3, sedangkan *cluster* 3 memiliki anggota paling sedikit diantara *cluster* 1 dan 2.

7. Validitas Metode K-Means

Berdasarkan hasil output nilai *Silhouette Index* yang diperoleh adalah sebesar 0,8641212 dan nilai *Partion Coefficient Index* yang diperoleh adalah sebesar 0,9684215. Maka *cluster* yang terbentuk dengan metode *K-means* valid, ditunjukkan pada nilai *Silhouette Index* dan *Partion Coefficient Index* yang mendekati nilai 1.

Clustering Menggunakan Metode Fuzzy C-means

1. Jumlah Cluster Terbentuk

Tabel 7. Hasil Cluster Metode FCM Yang Terbentuk

Cluster	1	2	3
Jumlah Anggota	34	2	4

Tabel 7 menunjukkan *cluster* yang terbentuk adalah sebanyak $k = 3$ dengan jumlah anggota masing-masing *cluster* yaitu 34, 2, 4 yang berarti hasil metode *Fuzzy C-means* dengan $k = 3$ terbentuk *cluster* 1 sebanyak 34 toko online, *cluster* 2 sebanyak 2 toko online, *cluster* 3 sebanyak 4 toko online.

2. Matriks Partisi Awal

Tabel 8. Matriks Partisi Awal

Cluster 1	Cluster 2	Cluster 3			
1 0.0123392663	0,09619027	0.0257579866	20 0.9919528899	0,0000786099	0.0072610115
2 0.1852956291	0,001073469	0.7973574821	21 0.9920096421	0,0000882062	0.0071082956
3 0.9119381773	0,000686020	0.0812016281	22 0.9963141480	0,0000423804	0.0032620477
4 0.0025364509	0,000072623	0.9967373228	23 0.9986770172	0,0000147872	0.0011751112
5 0.0387177763	0,002265589	0.9386263307	24 0.9954774669	0,0000523676	0.0039988576
6 0.9976878193	0,000023623	0.0020759544	25 0.9989479424	0,0000117283	0.0009347748
7 0.9897410557	0,000099155	0.0092673989	26 0.9970920485	0,0000331697	0.025762546
8 0.9832480275	0,000156729	0.0151846800	27 0.0257459210	0,09015151	0.0727389421
9 0.0008762888	0,000023461	0.9988890990	28 0.9819666553	0,0001683298	0.0163500471
10 0.9575588270	0,0003658978	0.0387821952	29 0.9969716657	0,0000342182	0.0026861525
11 0.9955030886	0,0000520617	0.0039762943	30 0.9988871333	0,0000124017	0.0009888493
Cluster 1	Cluster 2	Cluster 3	31 0.9955980448	0,0000508114	0.0038938417
12 0.9996522423	0,0000038079	0.0003096792	32 0.9964801232	0,0000403884	0.0031159925
13 0.9801267344	0,0001840443	0.0180328222	33 0.9964277944	0,0000361769	0.0032104367
14 0.9993400102	0,000006.967	0.0005903208	34 0.9982403685	0,0000197797	0.0015618342
15 0.9971899350	0,0000320182	0.0024898830	35 0.9985629927	0,0000160747	0.0012762604
16 0.9956990446	0,0000497119	0.0038038357	36 0.9973133166	0,0000305712	0.0023809718
17 0.9956625437	0,0000501535	0.0038359209	37 0.9959700637	0,0000464769	0.0035651673
18 0.9992334655	0,0000084169	0.0006823656	38 0.9987006284	0,0000145153	0.0011542187
19 0.9934904894	0,0000643034	0.0058664762	39 0.9990872824	0,0000101297	0.0008114209
			40 0.9946790534	0,0000613483	0.0047074631

Tabel 8 menunjukkan matriks partisi dari masing-masing toko online terhadap cluster 1, 2 dan 3

3. Pusat Cluster

Tabel 9. Pusat Cluster

	Twitter	Instagram	Facebook
Cluster 1	27245.86	113645.1	449655.4
Cluster 2	197068.91	390515.1	20250724.4
Cluster 3	188835.60	371635.1	7318161.3

Tabel 9 menunjukkan pusat *cluster* yang terbentuk dari masing-masing *cluster* terhadap ketiga variabel yaitu *cluster* 1 (27245.86, 113645.1, 449655.4), *cluster* 2 (197068.91, 390515.1, 20250724.4), dan *cluster* 3 (188835.60, 371635.1, 7318161.3).

4. Fungsi Obyektif

Nilai fungsi obyektif digunakan untuk menentukan anggota masuk kedalam *cluster* 1, 2 atau 3

Tabel 10. Jarak Anggota Terhadap Pusat Cluster

No	cluster 1	cluster 2	cluster3	20	2.94×10^{11}	3.71×10^{14}	4.02×10^{13}
1	4.98×10^{14}	6.39×10^{12}	2.39×10^{14}	21	3.54×10^{11}	3.98×10^{14}	4.94×10^{13}
2	2.15×10^{13}	2.30×10^{14}	5.01×10^{12}	22	1.74×10^{11}	4.09×10^{14}	5.31×10^{13}
3	2.50×10^{12}	3.32×10^{14}	2.81×10^{13}	23	5.94×10^{10}	4.01×10^{14}	5.05×10^{13}
4	4.77×10^{13}	1.67×10^{14}	1.21×10^{11}	24	2.16×10^{11}	4.10×10^{14}	5.37×10^{13}
5	7.39×10^{13}	1.26×10^{14}	3.05×10^{12}	25	4.70×10^{10}	4.01×10^{14}	5.03×10^{13}
6	9.01×10^{10}	3.81×10^{14}	4.33×10^{13}	26	1.35×10^{11}	4.06×10^{14}	5.23×10^{13}
7	3.70×10^{11}	3.69×10^{14}	3.95×10^{13}	27	2.87×10^{14}	8.20×10^{12}	1.02×10^{14}
8	5.78×10^{11}	3.63×10^{14}	3.75×10^{13}	28	6.21×10^{11}	3.63×10^{14}	3.73×10^{13}
9	4.56×10^{13}	1.70×10^{14}	4.00×10^{10}	29	1.39×10^{11}	4.04×10^{14}	5.15×10^{13}
10	1.33×10^{12}	3.48×10^{14}	3.29×10^{13}	30	4.97×10^{10}	4.01×10^{14}	5.02×10^{13}
11	2.15×10^{11}	4.10×10^{14}	5.37×10^{13}	31	2.09×10^{11}	4.10×10^{14}	5.34×10^{13}
12	1.51×10^{10}	3.96×10^{14}	4.87×10^{13}	32	1.65×10^{11}	4.08×10^{14}	5.29×10^{13}
13	6.77×10^{11}	3.61×10^{14}	3.68×10^{13}	33	1.38×10^{11}	3.79×10^{14}	4.27×10^{13}
14	2.70×10^{10}	3.88×10^{14}	4.57×10^{13}	34	7.98×10^{10}	4.03×10^{14}	5.10×10^{13}
15	1.30×10^{11}	4.06×10^{14}	5.22×10^{13}	35	6.46×10^{10}	4.02×10^{14}	5.06×10^{13}
16	2.05×10^{11}	4.10×10^{14}	5.36×10^{13}	36	1.24×10^{11}	4.06×10^{14}	5.21×10^{13}
No	cluster 1	cluster 2	cluster3	37	1.91×10^{11}	4.09×10^{14}	5.34×10^{13}
17	2.06×10^{11}	4.10×10^{14}	5.36×10^{13}	38	5.83×10^{10}	4.01×10^{14}	5.04×10^{13}
18	3.35×10^{10}	3.97×10^{14}	4.90×10^{13}	39	4.05×10^{10}	4.00×10^{14}	4.99×10^{13}
19	2.42×10^{11}	3.73×10^{14}	4.09×10^{13}	40	2.53×10^{11}	4.10×10^{14}	5.34×10^{13}

Tabel 10 menunjukkan jarak masing-masing anggota terhadap *cluster* 1,2 dan 3 berdasarkan perhitungan fungsi obyektif anggota *cluster* 1 berjumlah 34 anggota, *cluster* 2 berjumlah 2 anggota, dan *cluster* 3 berjumlah 4 anggota.

5. Hasil Proses Pengelompokan Anggota

Tabel 11. Hasil Anggota Cluster Metode FCM

Fuzzy c-means	
Cluster 1	Bukalapak, JD.ID, Bhineka, Eleven, Matahari, Aliexpress, Orami, Blanja, Qoo10, Jakmall, Sociolla, Asmara, Otten_Coffe, Alfacart, Berrybenka, Hijub, Dinomarket, Mapemall, Weshop, Bobobobo, Vip_plaza, Bro.do, Mothercare, Orori, Bukupedia, Tees, Hijabenka, Elektronik_solusion, Elektronik_city, Maskoolin, Pemmz, Muslimarket, Mamaway, dan 8Wood.
Cluster 2	Lazada dan Shepora
Cluster 3	Tokopedia, Blibli, Shope, dan Zalora

Cluster 1 beranggotakan 34 anggota dengan karakteristik jumlah *followers* sedikit.
 Cluster 2 beranggotakan 2 anggota dengan karakteristik jumlah *followers* banyak.
 Cluster 3 beranggotakan 4 anggota dengan karakteristik jumlah *followers* sedang.

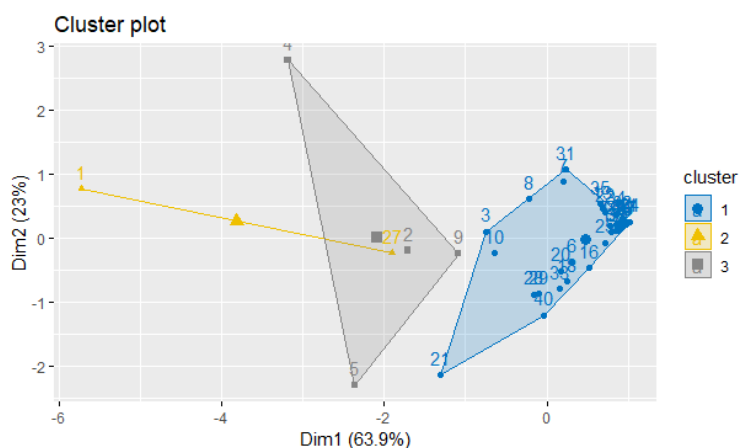
6. Derajad Keanggotaan

Digunakan untuk melihat pengelompokan anggota cluster.

Tabel 12. Derajad Keanggotaan

No	Toko Online	Cluster 1	Cluster 2	Cluster 3
1	Lazada	0.01	0.96	0.03
2	Tokopedia	0.19	0.02	0.80
3	Bukalapak	0.91	0.01	0.08
4	Blibli	0.00	0.00	1.00
5	Shope	0.04	0.02	0.94
6	JD ID	1.00	0.00	0.00
7	Bhineka	0.99	0.00	0.01
8	Elevenia	0.98	0.00	0.02
9	Zalora	0.00	0.00	1.00
10	Mataharimall	0.96	0.00	0.04
11	AliExpress	1.00	0.00	0.00
12	Orami	1.00	0.00	0.00
13	Blanja	0.98	0.00	0.02
14	Qoo10	1.00	0.00	0.00
15	Jakmall	1.00	0.00	0.00
16	Sociolla	1.00	0.00	0.00
17	Asmaraku	1.00	0.00	0.00
18	Otten Coffee	1.00	0.00	0.00
19	Alfacart	0.99	0.00	0.01
20	Berrybenka	0.99	0.00	0.01
21	Hijub	0.99	0.00	0.01
22	Dinomarket	1.00	0.00	0.00
23	Mapemall	1.00	0.00	0.00
24	Weshop	1.00	0.00	0.00
25	Bobobobo	1.00	0.00	0.00
26	Vip Plaza	1.00	0.00	0.00
27	Sephora	0.03	0.90	0.07
28	Bro.do	0.98	0.00	0.02
29	Mothercare	1.00	0.00	0.00
30	Orori	1.00	0.00	0.00
31	Bukupedia	1.00	0.00	0.00
32	Tees	1.00	0.00	0.00
33	Hijabenka	1.00	0.00	0.00
34	E.Solusion	1.00	0.00	0.00
35	Elektronik City	1.00	0.00	0.00
36	Maskoolin	1.00	0.00	0.00
37	Pemmz	1.00	0.00	0.00
38	Muslimarket	1.00	0.00	0.00
39	Mamaway	1.00	0.00	0.00
40	8Wood	0.99	0.00	0.00

7. Plot Cluster FCM



Gambar 9. Plot FCM

Gambar 9 menunjukkan pola persebaran setiap anggota yang terbentuk. Pada *cluster* 1 adalah yang paling banyak diantara *cluster* 2 dan 3, sedangkan *cluster* 2 memiliki anggota paling sedikit diantara *cluster* 1 dan 3.

8. Validitas Fuzzy C-means

Berdasarkan hasil output nilai *Silhouette Index* yang diperoleh adalah sebesar 0,9527799 dan nilai *Partion Coefficient Index* yang diperoleh adalah sebesar 0,9685199. Maka *cluster* yang terbentuk dengan metode *Fuzzy C-means* valid, ditunjukkan pada nilai *Silhouette* dan *Partion Coefficient Index Fuzzy C-means* mendekati nilai 1.

9. Perbandingan Metode K-Means dan Fuzzy C-means

Kriteria nilai yang digunakan dalam menentukan metode terbaik adalah dengan menggunakan nilai *Silhouette Index* dan *Partion Coefficient Index*.

Tabel 13. Output Nilai SI dan PCI

NILAI	METODE	
	K-Means	Fuzzy c-means
Silhouette Index (SI)	0,86	0,9527
Partion Coefficient Index (PCI)	0,9684	0,9685

Tabel 13 menunjukkan bahwa nilai *Silhouette Index* dan *Partion Coefficient Index* pada metode *Fuzzy c-means* lebih besar dibandingkan dengan nilai *Silhouette Index* dan *Partion Coefficient Index* pada metode K-means, sehingga dapat disimpulkan bahwa metode yang terbaik dalam pengklasteran segmentasi *e-commerce* di Indonesia 1 Juni 2018 adalah metode *Fuzzy C-means*.

4. KESIMPULAN

- 1) Gambaran segmentasi *e-commerce* di Indonesia pada 1 Juni 2018 menunjukkan rata-rata jumlah *followers* paling banyak Lazada 7896533 *followers*. Rata-rata *followers* paling rendah Weshop 100 *followers*. Kemudian jumlah *followers* paling banyak Lazada 22768300 *followers*, sedangkan jumlah *followers* paling sedikit Weshop, Blanja dan Qoo10 yaitu sebesar 0 *followers*.
- 2) Hasil dari penentuan jumlah cluster optimal adalah sebagai berikut:
 1. Metode elbow (gambar 5) diperoleh jumlah *cluster* optimal yaitu 3 *cluster*.
 2. Metode silhouette (gambar 6) diperoleh jumlah *cluster* optimal adalah 3 *cluster*.
 3. Metode gap statistic (gambar 7) diperoleh jumlah *cluster* optimal yaitu 3 *cluster*.
- 3) Hasil clustering menggunakan metode *K-means* dan *Fuzzy c-means* adalah sebagai berikut:
 1. Metode *K-means* jumlah *cluster* yang terbentuk adalah 3 *cluster*, jumlah anggota *cluster* 1 sebesar 34 toko online dengan karakteristik rata-rata jumlah *followers* sedang, *cluster* 2 sebesar 4 toko online dengan karakteristik rata-rata jumlah *followers* sedikit dan *cluster* 3 sebesar 2 toko online dengan karakteristik rata-rata jumlah *followers* banyak
 2. Metode *Fuzzy c-means* jumlah *cluster* yang terbentuk adalah 3 *cluster*, jumlah anggota *cluster* 1 sebesar 34 toko online dengan karakteristik rata-rata jumlah *followers* sedikit, *cluster* 2 sebesar 2 toko online dengan karakteristik rata-rata jumlah *followers* banyak dan *cluster* 3 sebesar 4 toko online dengan karakteristik rata-rata jumlah *followers* sedang.
- 4) Hasil perbandingan metode *K-means* dan *Fuzzy c-means* yang terbaik dalam pengklasteran segmentasi *e-commerce* adalah metode *Fuzzy c-means* karena nilai SI dan PCI *Fuzzy c-means* lebih besar dari pada nilai SI dan PCI *K-Means*.

Ucapan Terima Kasih

Dalam penyusunan tulisan ini, banyak pihak yang telah memberikan dukungan kepada penulis. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih kepada seluruh dosen dan pimpinan Jurusan Statistika Institut Sains & Teknologi AKPRIND Yogyakarta.

Daftar Pustaka

- Bezdek JC. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York. Plenum Press.
- Fuady M. (2005). *Pengantar Hukum Bisnis: Menata Bisnis Modern di Era Globalisasi*. Bandung: PT Citra Aditya Bakti.
- Jatipaningrum MT. (2011). *Fuzzy Klustering Short Time Series Untuk Segmentasi Pelanggan (Studi Kasus: Sampel Loads Pelanggan PLN di Yogyakarta)*. Yogyakarta: Universitas Gadjah Mada.
- Ramadhan A. (2017). *Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling*. Pekanbaru Riau: Universitas Islam Negeri Sultan Syarif Kasim.
- Selviana N. (2016). *Analisis Perbandingan K-Means dan Fuzzy C-Means untuk Pemetaan Motivasi Balajar Mahasiswa*. Pekanbaru Riau: Universitas Islam Negeri Sultan Syarif Kasim.
- Suryowati K dan Harmastuti. (2013). *Aljabar Linier*. Yogyakarta. Akprind Pres.
- Suryowati K, dkk. (2013). *Statistik Deskriptif*. Yogyakarta. Akprind Pres.