

## PERBANDINGAN PENGELOMPOKAN USAHA MIKRO KECIL DAN MENENGAH DI KABUPATEN KLATEN TAHUN 2019 DENGAN METODE K-MEANS DAN CLUSTERING LARGE APPLICATION

Titik Zulyanti<sup>1</sup>, Noeryanti<sup>2\*</sup>

Jurusan Statistika, Institut Sains & Teknologi AKPRIND Yogyakarta

Email: [titikzulyanti@gmail.com](mailto:titikzulyanti@gmail.com)<sup>1</sup>, [snoeryanti@yahoo.com](mailto:snoeryanti@yahoo.com)<sup>2</sup>

\*Corresponding Author

### *Abstract*

Micro, Small and Medium Business or MSMBs have been proven to provide employment and become the driving force of the economy in Indonesia. On the other hand, MSMEs also face various problems, so it is important to know the characteristics of a group in order to determine strategies as an effort to optimize results. This study aims to determine the characteristics of MSMBs in Klaten Regency based on turnover, assets, number of workers and length of business. The results of segmentation will divide MSMBs into several clusters where each cluster has similar or almost the same business characteristics. The method used to perform this analysis is the K-Means Method and Clustering Large Application. The results of the analysis show that the best method is Clustering with K-Means, because it has a higher Silhouette Index value, which is 0.78. The number of clusters formed is 2, with the number of cluster 1 or groups of small entrepreneurs as much as 6711 and cluster 2 or groups of large entrepreneurs as much as 185. The number of small entrepreneurs or MSMBs who are less productive is expected to be a concern in solving problems or designing policies so that they can increase the productivity of MSMBs in Klaten Regency, Indonesia.

**Keywords:** MSMBs, *cluster analysis*, *K-Means*, *clustering large application*

### **Abstrak**

Usaha Mikro Kecil dan Menengah atau UMKM telah terbukti memberikan lapangan kerja dan menjadi penggerak roda perekonomian di Indonesia. Di sisi lain, UMKM juga menghadapi berbagai masalah, sehingga penting untuk mengetahui karakteristik dari suatu kelompok agar dapat menentukan strategi sebagai usaha untuk mengoptimalkan hasil. Penelitian ini bertujuan untuk mengetahui karakteristik dari UMKM di Kabupaten Klaten berdasarkan omzet, aset, jumlah tenaga kerja serta lamanya usaha. Hasil segmentasi akan membagi UMKM ke dalam beberapa klaster dimana masing-masing klaster memiliki karakteristik usaha yang mirip atau hampir sama. Metode yang digunakan untuk melakukan analisis ini adalah Metode *K-Means* dan *Clustering Large Application*. Hasil analisis menunjukkan bahwa metode terbaik adalah Clustering dengan K-Means, karena memiliki nilai *Silhouette Index* yang lebih tinggi, yaitu 0,78. Jumlah *cluster* yang terbentuk adalah 2, dengan jumlah *cluster* 1 atau kelompok pengusaha kecil sebanyak 6711 dan *cluster* 2 atau kelompok pengusaha besar sebanyak 185. Banyaknya pengusaha kecil atau UMKM yang kurang produktif diharapkan menjadi perhatian dalam penyelesaian masalah atau perancangan kebijakan sehingga dapat meningkatkan produktifitas UMKM di Kabupaten Klaten.

**Kata Kunci :** UMKM, *analisis cluster*, *K-Means*, *clustering large application*

### **1. Pendahuluan**

Usaha Mikro, Kecil, dan Menengah (UMKM) dipandang dapat menstabilkan perekonomian nasional. Oleh sebab itu, UMKM perlu diberdayakan sebagai bagian integral ekonomi rakyat yang mempunyai kedudukan, peran, dan potensi strategis untuk mewujudkan struktur perekonomian nasional yang makin seimbang, berkembang, dan berkeadilan. Dalam mewujudkan harapan tersebut tentu membutuhkan suatu strategi, salah satu cara yang dapat dilakukan adalah dengan menyusun kebijakan yang berdasarkan data yang sesuai dengan keadaan lapangan sehingga dapat diperoleh hasil analisis mengenai karakteristik dari suatu kelompok.

Penelitian ini dilakukan untuk melakukan pengelompokan UMKM di Kabupaten Klaten Tahun 2019 berdasarkan omzet, aset, jumlah tenaga kerja serta lamanya usaha. Hasil segmentasi akan membagi UMKM ke dalam beberapa klaster dimana masing-masing klaster memiliki karakteristik Usaha yang mirip atau hampir sama. Metode yang digunakan untuk melakukan pengelompokan adalah Metode *K-Means* dan *Clustering Large Application*. Dari kedua metode ini nantinya akan dipilih metode terbaik sebagai acuan dalam penyelesaian masalah atau perancangan kebijakan dalam meningkatkan produktifitas UMKM di Kabupaten Klaten.

Penelitian terdahulu terkait segmentasi dan penerapan metode *K-Means* dan *Clustering Large Application* antara lain:

- a. penelitian yang dilakukan oleh Athifaturrofifah, dkk (2019) dengan judul “Perbandingan Pengelompokan *K-Means* dan *K-Medoids* Pada Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Studi Kasus: Data Titik Panas di Indonesia pada 28 April 2018)”. Hasil penelitian menunjukkan bahwa metode *K-Means* dapat memberikan hasil pengelompokan yang lebih baik karena memiliki nilai *Silhouette Coefficient* atau nilai *SC* lebih besar dari pada *K-Medoids*.
- b. penelitian lain yang dilakukan oleh Sangga V.A.P (2018) dengan judul “Perbandingan Algoritma *K-Means* dan Algoritma *K-Medoids* dalam Pengelompokan Komoditas Peternakan di Provinsi Jawa Tengah”. Hasil penelitian menunjukkan bahwa pengelompokan terbaik dengan perbandingan variansi cluster yaitu metode *K-Medoids* karena memiliki nilai variansi cluster terkecil.
- c. Penelitian yang dilakukan oleh Anggraeni N.L, dkk (2020) dengan judul “Komparasi Algoritma *K-Means* dan *K-Medoids* untuk Menangani Strategi Promosi di Politeknik TEDC Bandung”. Dari Penelitian ini dapat disimpulkan bahwa algoritma *K-Means* yang memiliki performansi lebih baik jika dibandingkan dengan algoritma *K-Medoids*.
- d. Penelitian lain yang dilakukan oleh Rifa I.H, dkk (2019) dengan judul “Implementasi Algoritma Clara Untuk Data Gempa Bumi di Indonesia”. Hasil penelitian menunjukkan bahwa jumlah cluster optimal berdasarkan magnitudo dan kedalaman terjadinya gempa adalah tiga cluster
- e. Penelitian yang dilakukan oleh Puntoriza, dkk(2020) dengan judul “Analisis Persebaran UMKM kota Malang Menggunakan *Cluster K-Means*”. Penelitian ini menghasilkan pengelompokan dengan cluster 1 yaitu cluster tinggi dan cluster 2 yaitu cluster sedang.

Berdasarkan penelitian-penelitian sebelumnya, maka akan dilakukan penelitian dengan metode *K-Means* dan *Clustering Large Application* untuk mengelompokkan UMKM di Kabupaten Klaten Tahun 2019. Perbedaan penelitian ini dengan penelitian-penelitian sebelumnya adalah penggunaan dataset atau sampel yang besar dengan perbandingan dua metode clustering. Hasil *clustering* dari kedua metode akan dibandingkan menggunakan *Silhouette Coefficient* dan kemudian dipilih metode terbaik. Dari kesimpulan yang diambil harapannya dapat membantu pemerintah dalam menentukan kebijakan sehingga dapat meningkatkan produktifitas UMKM serta meningkatkan kesejahteraan masyarakat khususnya di Kabupaten Klaten.

## 2. Metode Penelitian

Penelitian ini menggunakan data sekunder yang diperoleh berdasarkan dokumen dan informasi dari Dinas Perdagangan Koperasi dan UKM Kabupaten Klaten Provinsi Jawa Tengah pada bulan September sampai dengan November 2020 dengan objek penelitiannya adalah data 6896 pelaku UMKM yang berada di Kabupaten Klaten Jawa Tengah Tahun 2019. Adapun Variabel yang digunakan mencakup Lama usaha (Tahun), Omset(Rp), Aset(Rp), Serta Jumlah Tenaga Kerja (Orang).

### 2.1 Usaha Mikro Kecil dan Menengah

UU Nomor 20 Tahun 2008 menjelaskan bahwa Dunia Usaha adalah Usaha Mikro, Usaha Kecil, Usaha Menengah, dan Usaha Besar yang melakukan kegiatan ekonomi di Indonesia dan berdomisili di Indonesia. Usaha Mikro Kecil dan Menengah (UMKM) merupakan usaha produktif yang telah terbukti memberikan lapangan kerja dan menjadi penggerak roda perekonomian di Indonesia.

Berikut kriteria atau ketentuan umum Usaha Mikro, Usaha Kecil dan Usaha Menengah berdasarkan UU No. 20 Tahun 2008 yang disajikan pada **Tabel 2.1**, **Tabel 2.2**, dan **Tabel 2.3**.

**Tabel 2.1** Kriteria Usaha Mikro

Aset	$\leq$ Rp50.000.000
Omzet	$\leq$ Rp300.000.000

**Tabel 2.2** Kriteria Usaha Kecil

Aset	$>$ Rp50.000.000 s.d Rp500.000.000
Omzet	$>$ Rp300.000.000 s.d Rp2.500.000.000

**Tabel 2.3** Kriteria Usaha Menengah

Aset	$>$ Rp500.000.000 s.d Rp10.000.000.000
Omzet	$>$ Rp2.500.000.000 s.d Rp50.000.000.000

## 2.2 Statistik Deskriptif

Statistik deskriptif atau statistik deduktif adalah statistik yang berkenaan dengan metode atau cara mendeskripsikan, menggambarkan, menjabarkan, atau menguraikan data sehingga mudah dipahami. Statistik deskriptif mengacu pada bagaimana menata atau mengorganisasi data, menyajikan, dan menganalisis data. (Rohmad, dkk, 2015)

## 2.3 Analisis Cluster

Analisis *cluster* adalah salah satu teknik dari *analisis multivariat* yang mempunyai tujuan utama untuk mengelompokkan objek-objek yang berdasarkan kemiripan karakteristik yang dimilikinya. Adapun prosedur dari analisis *cluster* secara garis besar yaitu: (Alwi W, dkk, 2018)

### a. Standarisasi Data

Perlunya dilakukan proses standarisasi data dengan *transformasi* pada data asli sebelum dianalisis lebih lanjut karena variabel yang diteliti memiliki perbedaan ukuran satuan yang besar. Standarisasi dilakukan terhadap variabel yang relevan ke dalam bentuk *z-score*. Rumus perhitungan *Z* adalah sebagai berikut:

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

Keterangan:

$X$  = nilai data suatu variabel

$\sigma$  = standar deviasi data suatu variabel

$\mu$  = rata-rata data suatu variabel

### b. Deteksi Outlier

*Outliers* adalah data pengamatan dengan nilainya “*ekstrim*” yang muncul karena memiliki karakteristik yang unik atau terlihat sangat jauh berbeda dari observasi-observasi lainnya. Deteksi data yang *outlier* dapat dilakukan dengan melakukan standarisasi data (*z-score*) setelah itu melakukan pengecekan dengan *boxplot*. Metode ini merupakan metode yang paling umum digunakan untuk deteksi pencilan. Metode ini mempergunakan nilai kuartil dan jangkauan. Kuartil 1, 2, dan 3 akan membagi sebuah urutan data menjadi empat bagian. Jangkauan atau *Interquartile Range (IQR)* didefinisikan sebagai selisih kuartil 1 terhadap kuartil 3 ( $IQR = Q3 - Q1$ ). Data-data pencilan dapat ditentukan dengan melihat

nilai yang kurang dari  $1.5 \cdot IQR$  terhadap kuartil 1 dan nilai yang lebih dari  $1.5 \cdot IQR$  terhadap kuartil 3.

### c. Asumsi-asumsi Cluster

Terdapat dua asumsi yang harus dipenuhi dalam analisis *cluster* yaitu tidak terjadi *multikolinearitas* dan sampel mewakili populasi. Untuk mendeteksi ada atau tidaknya *multikolinearitas* dapat dilakukan dengan menghitung koefisien korelasi sederhana (*korelasi Pearson*) antar variabel bebas. Berikut ini adalah rumus *korelasi pearson*:

$$r = \frac{n \sum_{i=1}^n X_{i1} X_{i2} - (\sum_{i=1}^n X_{i1}) (\sum_{i=1}^n X_{i2})}{\sqrt{(n \sum_{i=1}^n X_{i1}^2 - (\sum_{i=1}^n X_{i1})^2) (n \sum_{i=1}^n X_{i2}^2 - (\sum_{i=1}^n X_{i2})^2)}} \quad (2)$$

Keterangan :

$r$  = koefisien korelasi *pearson* antara dua variabel  $X_1$  dan  $X_2$

$n$  = banyaknya pengamatan

$X_{i1}$  = nilai pengamatan ke- $i$  pada variabel  $X_1$

$X_{i2}$  = nilai pengamatan ke- $i$  pada variabel  $X_2$

$i$  = 1, 2, 3, ...,  $n$

jika terdapat nilai yang mencapai atau melebihi 0,8 maka terjadi *multikolinearitas*. Dalam mengatasi masalah *multikolinearitas* pada data adalah dengan Metode *Principal Component Analysis (PCA)* atau analisis komponen utama. Selanjutnya untuk mengetahui apakah sampel yang diambil benar-benar dapat mewakili populasi yang ada dibutuhkan nilai Kaiser-Meyer Olkin (KMO). Rumus KMO adalah sebagai berikut:

$$KMO = \frac{\sum_{j=1}^P \sum_{k=j+1}^P r_{X_j X_k}^2}{\sum_{j=1}^P \sum_{k=1}^P r_{X_j X_k}^2 + \sum_{j=1}^P \sum_{k=j+1}^P \rho_{X_j X_k X_1}^2} \quad (3)$$

Keterangan:

$P$  = Banyaknya variabel

$r_{X_j X_k}$  = Korelasi antar variabel

$X_j$  = rata-rata variabel  $X_j$

$X_k$  = rata-rata variabel  $X_k$

$n$  = banyaknya observasi (objek)

$\rho_{X_j X_k X_1}$  = korelasi parsial antara variabel  $X_j$  dan  $X_k$  dengan  $X_1$

Dimana nilai KMO kurang dari 0.5 menandakan bahwa sampel yang diambil tidak dapat mewakili populasi yang ada.

### d. Ukuran Kemiripan Objek

Dalam menjelaskan atau mengukur kemiripan/kesamaan antar objek digunakan pendekatan ukuran jarak (*distance*). Semakin besar nilai jarak antara objek berarti semakin besar pula perbedaan karakteristik antara objek. Sedangkan semakin kecil nilai jarak antara objek berarti semakin tinggi pula kemiripan/kesamaan karakteristik antara objek. Terdapat dua pengamatan dalam ruang  $p$ -dimensi  $x' = (x_1, x_2, \dots, x_n)$  dan  $y' = (y_1, y_2, \dots, y_n)$ . Jarak dari dua pengamatan tersebut dapat dihitung dalam berbagai cara, seperti *jarak Euclid*. Jika jarak dua titik  $x$  dan  $y$  ditulis dengan  $d(x, y)$ , rumus perhitungannya dapat dihitung sebagai berikut:

$$d(x, y) = \|x - y\| = \sqrt{\sum_{j=1}^m (x_{ij} - y_{ij})^2} \quad (4)$$

Keterangan:

Objek pengamatan =  $i = 1, 2, \dots, n$

Variabel =  $j = 1, 2, \dots, m$

### e. Metode Pengelompokan

Terdapat dua macam metode dalam proses analisis *cluster* yaitu metode *hirarki* dan metode *non hirarki*. Adapun metode *hierarki* adalah metode pengelompokan yang

terstruktur dan bertahap serta jumlah kelompok atau *cluster* belum diketahui. Sedangkan pada metode *non hirarki*, telah ditentukan jumlah kelompok terlebih dahulu.

Metode *non hierarki method* atau *partitioning method* dalam penentuan jumlah *cluster* yang optimal dapat dilakukan sebelum proses *clustering*. Pengambilan keputusan dalam penentuan banyaknya *cluster* yang akan dibentuk dapat digunakan kriteria *silhouette*. Kriteria *silhouette* dapat digunakan dengan melihat plot antara nilai *overall average silhouette width* dengan jumlah *cluster* yang terbentuk. Nilai *k* dengan *overall average silhouette width* tertinggi digunakan untuk analisis.

Penentuan jumlah *cluster* juga dapat dilakukan dengan metode *elbow*. Metode *elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat presentase hasil perbandingan antara jumlah *cluster* yang akan siku pada suatu titik. Metode *Elbow* ini memberikan ide atau gagasan dengan cara memilih nilai *cluster* dan kemudian menambah nilai *cluster* tersebut untuk dijadikan model data dalam penentuan *cluster* terbaik. Dan selain itu persentase perhitungan yang dihasilkan menjadi pemanding antara jumlah *cluster* yang ditambah. Hasil presentase 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. (Muningsih E, 2017)

Penentuan jumlah *cluster* yang optimal juga dapat dilakukan dengan metode *gap statistik*, dimana kriteria banyak *cluster optimum* diberikan oleh nilai *gap statistik (k)* yang paling tinggi, atau yang pertama kali mengindikasikan kenaikan *gap* yang minimum jika *gap* selalu naik. (Silvi R, 2018)

#### f. Interpretasi Cluster

Tahapan interpretasi *cluster* adalah untuk mencari karakter setiap kelompok yang khas, salah satunya dapat dilakukan dengan membandingkan *mean* atau *centroid* pada setiap kelompok

### 2.4 K-Means Clustering

*Clustering* adalah pembentukan kelompok dimana suatu objek ditempatkan dalam kelompok yang sama dengan objek yang lain apabila memiliki kemiripan karakteristik, namun menempati kelompok yang berbeda apabila tidak memiliki kemiripan atau berbeda karakteristik. Tujuan dari *clustering* adalah meminimalkan jarak di dalam *cluster* dan memaksimalkan jarak antar *cluster*. Sebuah *cluster* adalah suatu kumpulan data yang mirip dengan lainnya atau tidak mirip data pada kelompok lain dan disimpulkan bahwa *clustering* merupakan usaha untuk mengidentifikasi kelompok objek yang mirip untuk membantu menemukan pola penyebaran dan pola hubungan dalam sekumpulan data yang besar. *Clustering* disebut sebagai *unsupervised learning* yaitu suatu teknik data mining yang tidak memerlukan adanya data latih terlebih dahulu. (Rifa I.H, dkk, 2019)

### 2.5 Clustering Large Application

Clara (Kaufman dan Rousseeuw, 1990) merupakan salah satu macam pengelompokan data dengan *medoid* sebagai pusat klasternya. *Medoid* merupakan objek yang letaknya terpusat pada suatu klaster, atau dengan kata lain merupakan suatu objek yang merepresentasikan anggota pada suatu data dan memiliki rata-rata perbedaan (*dissimilarity*) yang paling kecil dengan anggota-anggota lain. Berbeda dengan metode *medoid* lainnya, yaitu *Pam*, metode *Clara* memiliki sifat *robust* terhadap *outlier* dan dapat digunakan untuk data dalam jumlah besar. Metode ini lebih efisien dalam hal waktu komputasi dan dalam penyimpanan data set yang besar. (Muslim, 2018)

### 2.6 Silhouette Coefficient

Metode *validasi silhouette index* merupakan salah satu ukuran *validasi* yang berbasis kriteria internal. *Silhouette index* akan mengevaluasi penempatan setiap objek dalam setiap kluster dengan membandingkan jarak rata-rata objek dalam satu kluster dan jarak antara objek dengan kluster yang berbeda (Aini et al., 2014). Menghitung *koefisien silhouette* yang didefinisikan sebagai rata-rata  $s(i)$  yaitu,

$$SI = \frac{1}{n} \sum_{i=1}^n s(i) \quad (8)$$

$$\text{dengan, } s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))}, b(i) = \min o(i, C), \text{ dan } a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} o(i, j)$$

$b(i)$  : nilai minimum dari jarak rata-rata objek  $i$  dengan semua objek pada kluster lain  $C$

$a(i)$  : rata-rata jarak objek ke- $i$  dengan semua objek yang berada di dalam satu kluster  $A$

Hasil perhitungan nilai *koefisien silhouette* berada pada range -1 sampai 1. Semakin besar nilai *koefisien silhouette* akan semakin baik kualitas suatu kelompok.

Hasil validasi digunakan untuk profilisasi hasil analisis kluster yang dilakukan pada hasil pengklusteran terbaik. Pada tahap profilisasi akan dilihat karakteristik dari tiap kluster yang terbentuk, sehingga dapat dilihat kecenderungan dari tiap kluster. (Nahdliyah M.A, dkk, 2019).

## 2.7 R-Studio

Aplikasi R menggunakan bahasa pemrograman R yang mana bahasa pemrograman ini berbasis pada bahasa pemrograman S. R memiliki kemampuan menganalisis data dengan sangat efektif dan dilengkapi dengan operator pengolahan array serta matriks. Kelebihan lain dari aplikasi R, ukuran file yang disimpan jauh lebih kecil dibanding software lainnya, lengkap dalam operator perhitungan array, dan terdiri dari koleksi tools statistik yang terintegrasi untuk analisis data, dapat dikembangkan sesuai kebutuhan dan sifatnya yang terbuka, setiap orang dapat menambahkan fitur-fitur tambahan dalam bentuk paket ke dalam software R dan dapat dipasang pada sistem operasi Windows, Mac OS, Mac OS X, Linux, Free BSD, NetBSD, irix, Solaris, AIX, dan HPUX. (Sihombing R.E, dkk, 2019)

Analisis klusterisasi menggunakan R-Studio dengan serangkaian *Package* yang digunakan meliputi *library tidyverse* yang digunakan untuk melakukan manipulasi data, *kemudian library cluster* yang digunakan untuk melakukan klusterisasi dan ada *package* tambahan untuk melakukan klusterisasi dan membuat visualisasi (grafik) yang menarik yaitu *library factoextra*.

## 3. Hasil dan Pembahasan

### 3.1 Statistik Deskriptif

Analisis Deskriptif digunakan untuk mendeskripsikan atau mengetahui gambaran mengenai karakteristik dari pelaku UMKM di Kabupaten Klaten berdasarkan Variabel Lama Usaha dengan satuan tahun, Omzet dan Aset dengan satuan rupiah, dan Tenaga Kerja dengan satuan jumlah orang atau jumlah tenaga kerja. Analisis deskriptif ini disajikan dalam bentuk ringkasan statistik yang meliputi jumlah data penelitian, nilai minimum, nilai maksimum, mean atau rata-rata dan penyebaran data atau *standar deviasi* pada masing-masing variabel penelitian. Analisis deskriptif dilakukan dengan *Software R dengan Package library pastest*, sebagaimana hasilnya disajikan pada **Tabel 3.1**.

**Tabel 3.1** Statistik Deskriptif

Variabel	N	Min	Maks	Mean	Std. Dev
Lama Usaha	6.896	0	75	11,11	9,78
Omset	6.896	175.000	4.000.000.000	60.540.000	172.062.559
Asset	6.896	75.000	1.100.000.000	17.080.000	51.283.274

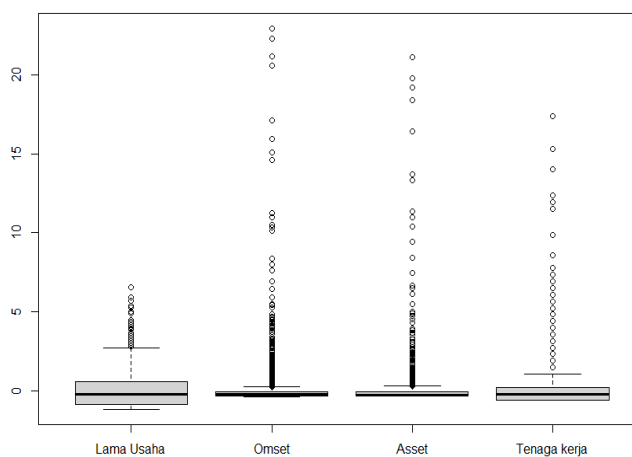
Tenaga Kerja	6.896	1	44	2,43	2,39
--------------	-------	---	----	------	------

Hasil analisis statistik deskriptif telah memberikan informasi mengenai karakteristik data Pelaku UMKM Kabupaten Klaten Tahun 2019. Namun, karena statistik deskriptif memberikan informasi hanya mengenai data yang dipunyai dan sama sekali tidak menarik kesimpulan yang lebih banyak dan lebih jauh maka akan dilakukan analisis *clustering* agar dapat menghasilkan informasi yang dapat dipergunakan lebih luas lagi.

### 3.2 Analisis Clustering

#### a. Standarisasi dan Deteksi Outlier

Standarisasi yang digunakan pada penelitian ini adalah konversi setiap variabel terhadap nilai standar (*Z Score*) dengan rumus (1). Setelah dilakukan Standarisasi selanjutnya dibuat boxplot untuk dianalisis apakah terdapat data outlier atau tidak. Berikut visualisasi untuk Boxplotnya.



**Gambar 3.1** Boxplot deteksi outlier

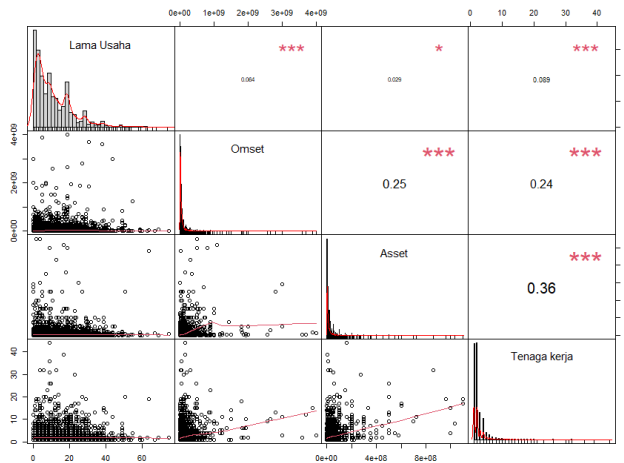
Dalam analisis *cluster*, data *outlier* bisa di pertahankan dan bisa juga dihilangkan sesuai kebutuhan peneliti. Namun, untuk analisis ini data *outlier* tidak dihilangkan karena terdiri dari beberapa variabel, sehingga jika dalam menghilangkan *outlier* difokuskan ke salah satu variabel, maka akan mengakibatkan outlier di variabel lain.

#### b. Uji Asumsi Clustering

Berdasarkan hasil analisis menggunakan *Software R-Studio* diperoleh hasil nilai *korelasi* antar variabel sebagaimana disajikan pada **Tabel 3.2** dan visualisasinya pada **Gambar 3.2**.

**Tabel 3.2** Nilai Korelasi antar Variabel

Variabel	Lama Usaha	Omset	Asset	Tenaga Kerja
Lama Usaha	1	0,06	0,02	0,09
Omset	0,06	1	0,25	0,24
Asset	0,02	0,25	1	0,35
Tenaga Kerja	0,09	0,24	0,35	1



Gambar 3.2 Plot Korelasi antar Variabel

Berdasarkan Tabel 3.2 dan Gambar 3.2. dapat dilihat nilai korelasi antar variabel penelitian dimana nilai koefisien korelasi yang dihasilkan tidak ada yang melebihi 0,8. Artinya tidak terdapat multikolinearitas pada variabel-variabel penelitian mengenai pelaku UMKM di Kabupaten Klaten, dengan demikian maka analisis selanjutnya memenuhi syarat untuk dilakukan uji lanjutan.

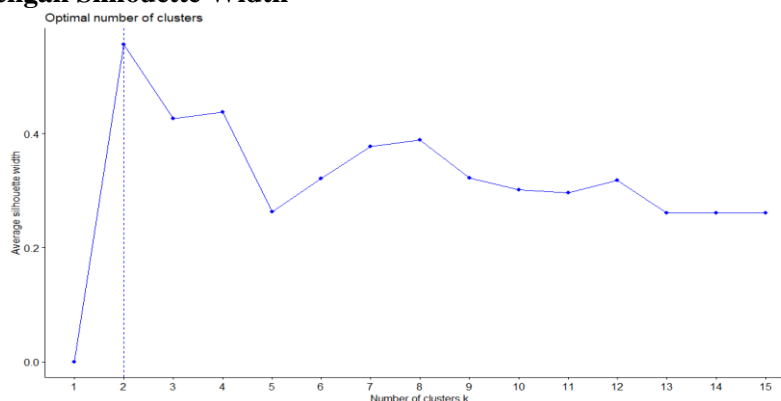
Sampel representative atau sampel yang mewakili dilakukan untuk menguji apakah data yang digunakan cukup untuk dianalisis lebih lanjut. Pada Uji KMO yang dilakukan dengan Software R-Studio dihasilkan hasil sebagai berikut:

```
> kmo(pkpi)
$KMO
[1] 0.6109142

$MSA
  Lama Usaha      Omset      Asset Tenaga kerja
0.6206815  0.6685176  0.5913157  0.5961307
```

Karena nilai  $KMO = 0,61 > 0,5$ , maka  $H_0$  tidak ditolak. Artinya dapat disimpulkan bahwa data yang digunakan layak untuk dilakukan analisis lanjut. Kemudian, dari analisis yang telah dilakukan untuk pengujian KMO, didapatkan juga nilai *Measure of Sampling Adequacy* (MSA). dapat diamati bahwa semua variabel menghasilkan  $MSA > 0,5$ . Artinya semua variabel penelitian ini masih dapat di prediksi dan layak untuk dianalisis lanjut.

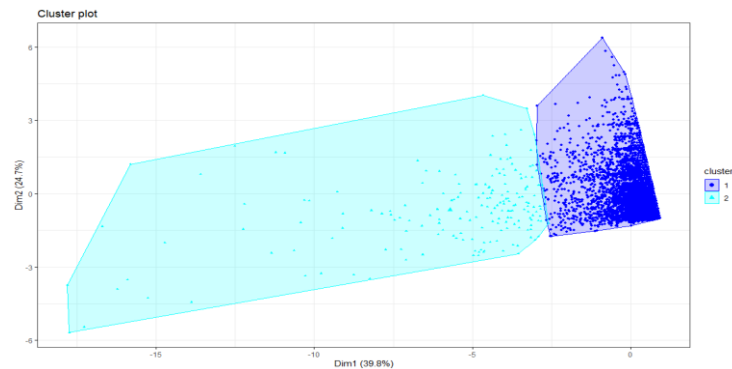
**3.3 K-Means Clustering**  
**a. K-Means dengan Silhouette Width**



Gambar 3.3 Grafik Sillhouette Width Metode K-Means



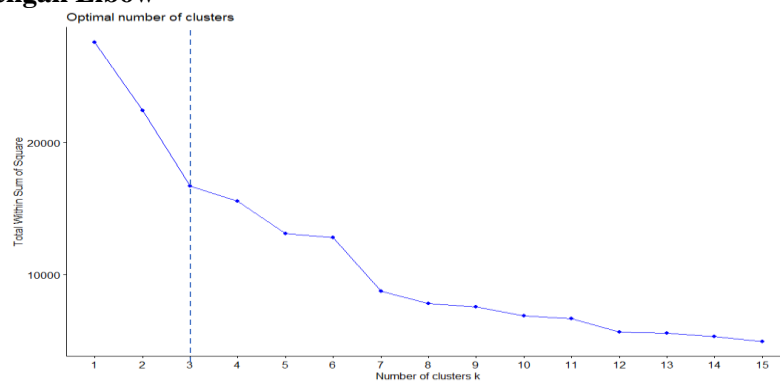
Berdasarkan **Gambar 3.3** dapat diamati bahwa nilai *Silhouette Width* yang paling tinggi adalah pada saat  $k=2$ . Oleh sebab itu, maka *clustering K-Means* akan dilakukan kedalam 2 kelompok. Berikut plot hasil analisisnya disajikan pada **Gambar 3.4**.



**Gambar 3.4** Plot Clustering K-Means  $k=2$

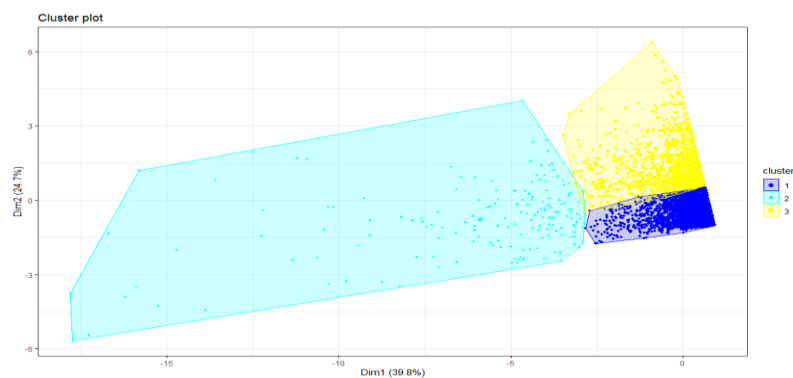
**Gambar 3.4** menunjukkan Plot data yang dikelompokkan dengan metode *K-Means* kedalam 2 *cluster*, dimana *cluster 1* ditunjukkan oleh warna biru tua dengan jumlah anggota kelompok sebanyak 6711, kemudian untuk *cluster 2* ditunjukkan oleh warna biru muda dengan jumlah anggota kelompok sebanyak 185.

#### b. K-Means dengan Elbow



**Gambar 3.5** Grafik Elbow Metode K-Means

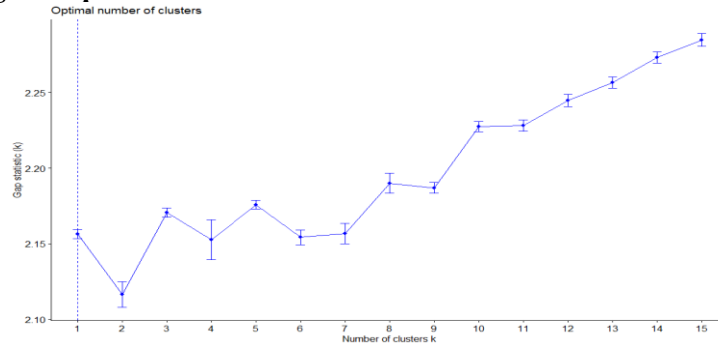
Berdasarkan **Gambar 3.3** dapat diamati bahwa garis yang mengalami patahan yang membentuk siku (*elbow*) adalah pada saat  $k=3$ . Oleh sebab itu, maka *clustering K-Means* akan dilakukan kedalam 3 kelompok. Berikut plot hasil analisisnya disajikan pada **Gambar 3.4**.



**Gambar 3.6** Plot Clustering K-Means  $k=3$

**Gambar 3.6** menunjukkan Plot data yang dikelompokkan dengan metode *K-Means* kedalam 3 cluster, dimana cluster 1 ditunjukkan oleh warna biru tua dengan jumlah anggota kelompok sebanyak 4716, sedangkan cluster 2 ditunjukkan oleh warna biru muda dengan jumlah anggota kelompok sebanyak 173, dan cluster 3 ditunjukkan oleh warna kuning dengan jumlah anggota kelompok sebanyak 2007.

**c. K-Means dengan Gap Statistik**

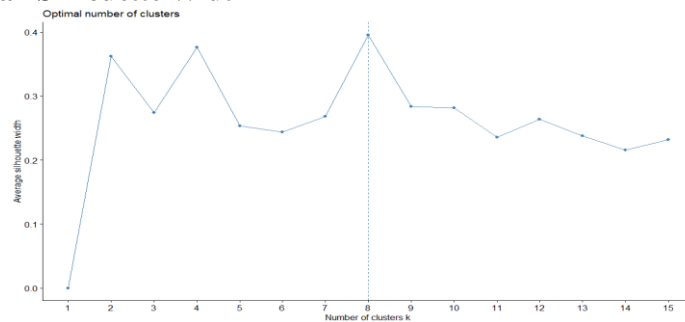


**Gambar 3.7** Grafik Gap Statistik Metode K-Means

Berdasarkan **Gambar 3.7** dapat dilihat bahwa banyak cluster optimum *K-Means* untuk metode gap statistik adalah  $k=1$  yaitu pada saat pertama kali mengindikasikan kenaikan. Kesimpulannya dengan metode *Gap Statistik* data tidak dapat dilakukan pengelompokan

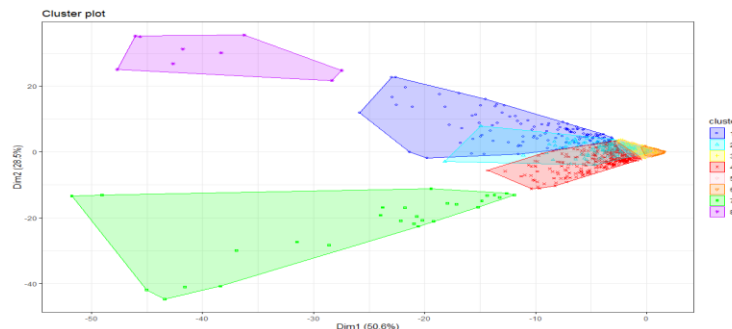
**3.4 Clustering Large Application**

**1. K-Means dengan Silhouette Width**



**Gambar 3.8** Grafik Silhouette Width Metode Clara

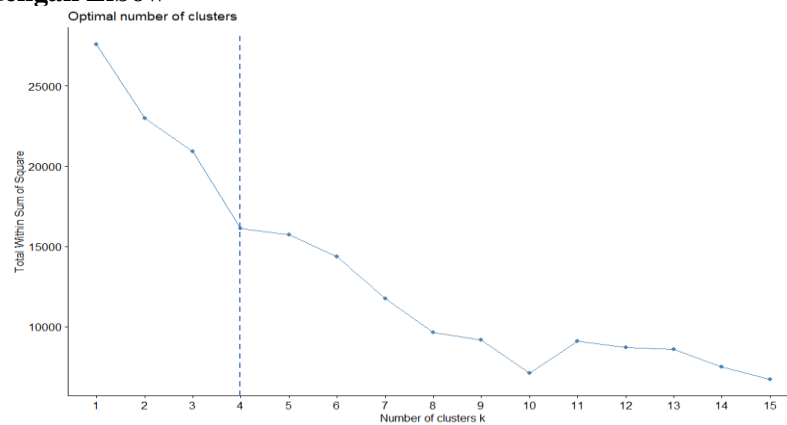
Berdasarkan **Gambar 3.8** dapat diamati bahwa nilai Silhouette Width yang paling tinggi adalah pada saat  $k=8$ . Oleh sebab itu, maka clustering *Clara* akan dilakukan kedalam 8 kelompok. Berikut plot hasil analisisnya disajikan pada **Gambar 3.9**.



**Gambar 3.9** Plot Clustering Clara  $k=8$

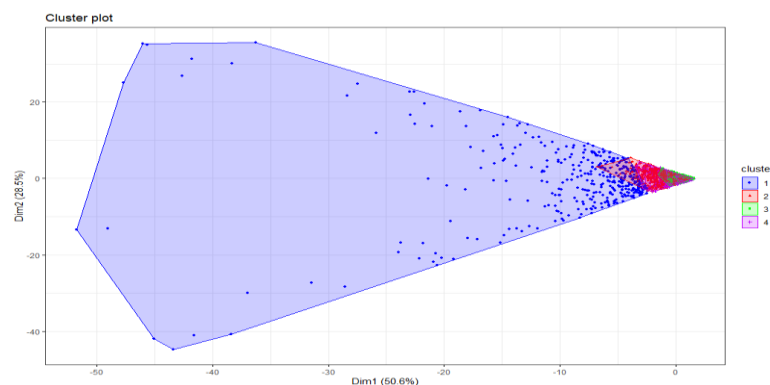
**Gambar 3.9** menunjukkan Plot data yang dikelompokkan dengan metode *Clara* kedalam 8 *cluster*, dimana *cluster* 1 ditunjukkan oleh warna biru tua dengan jumlah anggota kelompok sebanyak 124, kemudian untuk *cluster* 2 ditunjukkan oleh warna biru muda dengan jumlah anggota kelompok sebanyak 418, untuk *cluster* 3 ditunjukkan oleh warna kuning dengan jumlah anggota kelompok sebanyak 2304, *cluster* 4 berwarna merah dengan jumlah anggota kelompok terbanyak yaitu 406, sedangkan untuk *cluster* 5 berwarna pink dengan jumlah anggota kelompok sebanyak 2057, selanjutnya untuk *cluster* 6 berwarna orange dengan jumlah anggota kelompoknya adalah 1550, *cluster* 7 berwarna hijau jumlah anggota kelompoknya adalah 28, dan terakhir untuk *cluster* 8 dengan warna ungu dan memiliki jumlah anggota kelompok paling sedikit, yaitu 9.

## 2. K-Means dengan Elbow



**Gambar 3.10** Grafik Elbow Metode Clara

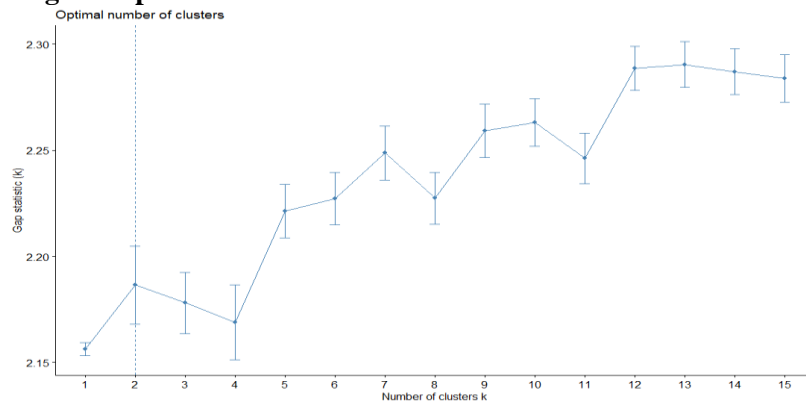
Berdasarkan **Gambar 3.10** dapat diamati bahwa garis yang mengalami patahan yang membentuk siku (*elbow*) adalah pada saat  $k=4$ . Oleh sebab itu, maka clustering *Clara* akan dilakukan kedalam 4 kelompok. Berikut plot hasil analisisnya disajikan pada **Gambar 3.11**.



**Gambar 3.11** Plot Clustering Clara  $k=4$

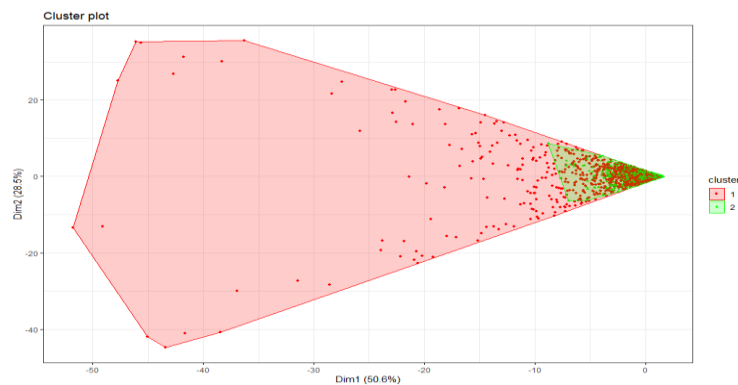
**Gambar 3.11** menunjukkan Plot data yang dikelompokkan dengan metode *Clara* kedalam data dengan metode *Clara* kedalam 4 *cluster*, dimana *cluster* 1 ditunjukkan oleh warna biru tua dengan jumlah anggota kelompok sebanyak 386, kemudian untuk *cluster* 2 ditunjukkan oleh warna merah dengan jumlah anggota kelompok sebanyak 2037, *cluster* 3 berwarna hijau dengan jumlah anggota kelompok 2636, dan terakhir untuk *cluster* 4 dengan warna ungu dan memiliki jumlah anggota kelompok paling sedikit, yaitu 1837.

### 3. K-Means dengan Gap Statistik



**Gambar 3.12** Grafik Gap Statistik Metode Clara

Berdasarkan **Gambar 3.12** dapat diamati bahwa nilai Silhouette Width yang paling tinggi adalah pada saat  $k=2$ . Oleh sebab itu, maka clustering K-Means akan dilakukan kedalam 2 kelompok. Berikut plot hasil analisisnya disajikan pada **Gambar 3.13**.



**Gambar 3.13** Plot Clustering Clara  $k=2$

**Gambar 3.11** menunjukkan Plot data yang dikelompokkan dengan metode Clara kedalam data dengan metode Clara kedalam kedalam 2 cluster, dimana cluster 1 ditunjukkan oleh warna merah dengan jumlah anggota kelompok sebanyak 2640, kemudian untuk cluster 2 ditunjukkan oleh warna hijau dengan jumlah anggota kelompok sebanyak 4256.

### 3.5 Perbandingan Hasil Clustering

Penentuan metode terbaik dilakukan dengan melihat nilai *silhouette index* terbesar dari hasil clustering menggunakan metode K-Means maupun Clara. Berdasarkan hasil analisis menggunakan software R-Studio (didapatkan hasil *Shilouette index* masing-masing hasil clustering seperti pada **Tabel 4..**

**Tabel 4.18** Nilai *Silhouette Index*

Metode	Cluster Optimum	Jumlah Cluster	<i>Silhouette Index</i>
K-Means	Silhouette	2	0,78
	Elbow	3	0,43
	Gap Stat	1	0
CLARA	Silhouette	8	0,22
	Elbow	4	0,28
	Gap Stat	2	0,37

Berdasarkan Tabel diatas dapat disimpulkan bahwa hasil *cluster* terbaik adalah dengan metode *K-Means* dengan jumlah *cluster* 2 karena memiliki nilai *Silhouette Index*nya lebih besar dibandingkan dengan yang lainnya. Selain itu hasil pengelompokan dengan Metode *K-Means* kedalam 2 *cluster* juga dapat dikatakan cukup baik karena memiliki nilai *Silhouette Index* di rentang 0,7-1

### 3.6 Profiling Cluster

*Profiling* dilakukan setelah diketahui hasil *clustering* terbaik, yaitu Clustering dengan metode *K-Means*. Berikut disajikan hasil analisis profiling cluster pada **Tabel 4.19**.

**Tabel 4.19** Profiling Cluster

Cluster	1	2
Lama Usaha (Th)	11,0	14,3
Omzet (Rp)	45247988	615367689
Aset (Rp)	1238010	18771378
Tenaga Kerja (orang)	2,21	10,6

Dari hasil analisis dapat dilihat bahwa nilai rata-rata setiap variabel pada cluster 2 memiliki rata-rata yang lebih tinggi dibandingkan dengan nilai rata-rata setiap variabel pada cluster 1. Dengan demikian dapat disimpulkan bahwa karakteristik atau ciri cluster dilihat dari nilai rata-ratanya cluster 1 termasuk kelompok pengusaha kecil, sedangkan cluster 2 merupakan kelompok pengusaha besar. Selain itu, jika dilihat dari karakteristik Usaha Mikro, Usaha Kecil Maupun Usaha menengah juga dapat disimpulkan bahwa cluster 1 terdiri dari Usaha Mikro dan Usaha Kecil, sedangkan untuk cluster 2 terdiri dari Usaha Menengah.

## 4. Kesimpulan

Hasil analisis statistik deskriptif telah memberikan informasi mengenai karakteristik data Pelaku UMKM Kabupaten Klaten Tahun 2019 dengan jumlah objek sebanyak 6896 dengan variabel Lama usaha, Omzet, Aset dan Tenaga kerja menunjukkan naik turun nya nilai atau terjadi fluktuasi pada setiap variabel.

Analisis *clustering* dilakukan dengan *standarisasi* untuk menyamakan skala pada masing-masing variabel, kemudian analisis mengindikasi adanya data *outlier*, namun tidak dihilangkan. Hasil analisis asumsi *cluster* yaitu tidak terjadi *multikolinearitas* dan sampel data yang di ambil sudah cukup mewakili populasi, sehingga dapat dilakukan analisis lanjutan.

Penentuan Jumlah *Cluster* optimum dilakukan dengan Metode *Silhouette Width*, *Elbow*, dan *gap statistic* dimana untuk metode *K-Means* dihasilkan cluster optimum pada saat  $k=2$ ,  $k=3$ , dan  $k=1$ . Sedangkan dengan metode *clara* dihasilkan cluster optimum pada saat  $k=8$ ,  $k=4$ , dan  $k=2$ . Perbandingan hasil *clustering* menunjukkan bahwa nilai *silhouette index* yang paling tinggi adalah pada hasil *clustering* *K-Means* dengan  $k=2$  yaitu sebesar 0,78.

*Clustering K-Means* dengan  $k=2$  pada data pelaku UMKM di Kabupaten Klaten Tahun 2019 terdiri dari *cluster* 1 dengan jumlah anggota kelompok sebanyak 6711 dengan karakteristik atau ciri *cluster* dilihat dari nilai rata-ratanya *cluster* 1 termasuk kelompok pengusaha kecil dan masuk dalam kategori usaha mikro dan kecil, sedangkan *cluster* 2 merupakan kelompok pengusaha besar dan masuk kategori usaha menengah dengan jumlah anggota kelompok sebanyak 185. Dari hasil ini, Pemerintah Kabupaten Klaten, khususnya Dinas Perdagangan Koperasi dan UKM sebaiknya memperbanyak kegiatan atau program dalam mengembangkan UMKM yang sudah ada, karena Sebagian besar pelaku UMKM yang ada di kabupaten Klaten masih termasuk Kelompok pengusaha kecil. Dengan demikian, diharapkan agar UMKM bisa lebih produktif serta dapat meningkatkan penyerapan tenaga kerja, meningkatkan pendapatan dan meningkatkan kesejahteraan

## Daftar Pustaka

- Alwi W, dkk, 2018, Analisis *Cluster* Untuk Pengelompokan Kabupaten/Kota Di Propinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat, Jurnal Msa Vol. 6 No. 1, Makassar.
- Anggraeini N.L, dkk, 2020, Komparasi *Algoritma K-Means* Dan *K-Medoids* Untuk Menangani Strategi Promosi Di Politeknik TEDC Bandung, TEDC Vol. 14 No.2, Bandung.
- Athifaturrofifah, dkk, 2019, Perbandingan Pengelompokan *K-Means* dan *K-Medoids* Pada Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Studi Kasus : Data Titik Panas Di Indonesia Pada 28 April 2018), Jurnal Eksponensial, Volume 10, Nomor 2, Samarinda.
- Berita Pembangunan, UMKM Berperan Mendukung Pencapaian Distribusi Pendapatan Guna Kurangi Kesenjangan, Kementerian PPN Bappenas, <https://www.bappenas.go.id/id/berita-dan-siaran-pers/umkm-berperan-mendukung-pencapaian-distribusi-pendapatan-guna-kurangi-kesenjangan/> (diakses tanggal 10 Oktober 2020)
- Berita Pembangunan, IDF 2019: Modernisasi Umkm Atasi Permasalahan Masyarakat, Bantu Perekonomian Nasional, Kementerian PPN Bappenas, <https://www.bappenas.go.id/id/berita-dan-siaran-pers/idf-2019-modernisasi-umkm-atasi-permasalahan-masyarakat-bantu-perekonomian-nasional/> (diakses tanggal 10 Oktober 2020).
- Ediyanto, dkk, 2013, Pengklasifikasian Karakteristik Dengan Metode *K-Means Cluster Analysis*, Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster) Volume 02, No. 2, Indonesia.
- Muningsih E, 2017, *Optimasi Jumlah Cluster K-Means Dengan Metode Elbow* Untuk Pemetaan Pelanggan, Prosiding Seminar Nasional Elinvo, Hal:105-114, Yogyakarta.
- Nahdliyah M.A, dkk, 2019, *Metode K-Medoids Clustering Dengan Validasi Silhouette Index dan C-Index*, Jurnal Gaussian, Volume 8, Nomor 2, Semarang.
- Pemerintah RI, 2008, Undang-Undang No. 20 Tahun 2008 Yang Mengatur Tentang USAHA MIKRO, KECIL, DAN MENENGAH, Lembaran Negara RI Tahun 2008 No. 20, Sekretariat Negara, Jakarta-Indonesia.
- Puntoriza, dkk, 2020, Analisis Persebaran UMKM Kota Malang Menggunakan *Cluster K-means*, Journal of Information System Vol. 5, No. 1, Salatiga.
- Rifa I.H, dkk, 2019, *Implementasi Algoritma Clara* Untuk Data Gempa Bumi Di Indonesia, Seminar Nasional Penelitian Pendidikan Matematika (SNP2M) 2019 UMT, Solo.
- Rohmad, dkk, 2015, Pengantar Statistika, Yogyakarta, KALIMEDIA.
- Sangga V.A.P, 2018, Perbandingan *Algoritma K-Means* dan *Algoritma K-Medoids* Dalam Pengelompokan Komoditas Peternakan Di Provinsi Jawa Tengah Tahun 2015, Tugas Akhir, FMIPA, Jurusan Statistika, Universitas Islam Indonesia, Yogyakarta.
- Setyawan Y, dkk, 2018, Statistika Dasar Dilengkapi dengan *software R*, Yogyakarta, AKPRIND PRESS.
- Sihombing R.E, dkk, 2019, Program Aplikasi Bahasa R Untuk Pengelompokan Objek Menggunakan *Metode K-Medoids Clustering*, Jurnal Eurekamatika, Bandung.
- Silvi R, 2018, Analisis *Cluster* dengan Data *Outlier* Menggunakan *Centroid Linkage* dan *K-Means Clustering* untuk Pengelompokan Indikator HIV/AIDS di Indonesia, Jurnal Matematika "MANTIK", Vol. 04 No. 01, Bandung.
- Susilowati B.K, dkk, 2018, Metode ROBPCA (*Robust Principal Component Analysis*) dan *Clara (Clustering Large Area)* pada Data dengan *Outlier* (Studi Kasus Data Laporan Indeks Kebahagiaan Dunia Tahun 2018), Jurnal Ilmu Komputer VOL. XIII, No 2, Yogyakarta.
- Waworuntu, M.N.V, Amin M.F, 2018, Penerapan Metode *K-Means* Pemetaan Calon Penerima Jamkesda, Kumpulan jurnaL Ilmu Komputer (KLIK) Volume 05, No.02 Universitas Lambung Mangkurat, Banjarbaru.