

## Kombinasi *Single Linkage* Dengan *K-Means Clustering* Untuk Pengelompokan Wilayah Desa Kabupaten Pemalang

### Combination *Single Linkage* With *K-Means Clustering* For Grouping Village Areas Of Pemalang District

Sintiya<sup>1</sup>, Tri Ginanjar Laksana<sup>2</sup>, Nia Annisa Ferani Tanjung<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Fakultas Informatika, Institut Teknologi Telkom Purwokerto  
17102141@ittelkom-pwt.ac.id <sup>1</sup>, anjarlaksana@ittelkom-pwt.ac.id <sup>2</sup>, nia@ittelkom-pwt.ac.id <sup>3</sup>

#### Abstrak

Metode *K-Means* sangat bergantung pada penentuan pusat *cluster* awal yang berdampak pada kualitas *cluster* yang dihasilkan, selain penentuan pusat *cluster* jumlah  $k$  yang akan digunakan juga dapat mempengaruhi kualitas *cluster* dari metode *K-Means*. Kemiskinan sebagian besar dialami oleh masyarakat pedesaan, hal ini dapat dilihat dari masih rendahnya fasilitas yang ada untuk melayani kepentingan masyarakat dalam berbagai bidang. Untuk menghindari ketimpangan yang terjadi diperlukan suatu rencana pembangunan yang sesuai dengan karakteristik kesejahteraan masyarakat di wilayah tersebut. Oleh karena itu, diperlukan sebuah upaya pengelompokan desa agar pengambilan kebijakan menjadi tepat sasaran. Salah satu algoritma *clustering* yang umum digunakan adalah algoritma *K-Means* karena cukup sederhana, mudah diimplementasikan, dan memiliki kemampuan untuk mengelompokkan data yang besar dengan sangat cepat. Namun algoritma *K-Means* memiliki kelemahan pada penentuan pusat *cluster* awal yang diberikan. Inisialisasi pusat *cluster* secara acak dapat memberikan hasil pembentukan *cluster* yang berubah-ubah (tidak konsisten). Untuk itu, metode *K-Means* perlu dikombinasikan dengan metode hirarki dalam penentuan pusat *cluster* awal. Metode kombinasi ini disebut sebagai *Hierarchical K-Means* yang merupakan penggabungan antara metode *hierarchical* dan *partitioning*, di mana proses *hierarchical* digunakan untuk mencari inisialisasi pusat *cluster* awal dan proses *partitioning* untuk mendapatkan *cluster* yang optimal. Metode hirarki yang digunakan pada penelitian ini adalah metode *single linkage*. Berdasarkan metode *Elbow* jumlah  $k$  yang direkomendasikan untuk penelitian ini yaitu  $k = 4$ . Kombinasi algoritma *single linkage* dan *k-means* dengan  $k = 4$  pada penelitian ini menghasilkan nilai *silhouette coefficient* sebesar 0.685 yang merupakan kategori klaster yang layak atau sesuai, sedangkan pengukuran evaluasi dengan *Davies Bouldin Index* menghasilkan nilai 0.577.

**Kata Kunci:** *Cluster, Davies Bouldin Index, K-means, Silhouette coefficient, Single linkage*

#### Abstract

*K-Means* is very dependent on determining the center *cluster* initial which has an impact on the quality of *clusters* resulting, in addition to determining the center of *cluster* the number of  $k$  that will be used it can also affect the quality of the cluster from the method *K-Means*. Poverty is mostly experienced by rural communities, this can be seen from the lack of existing facilities to serve the interests of the community in various fields. To avoid the imbalance that occurs, a development plan is needed in accordance with the characteristics of the welfare of the people in the region. Therefore, we need an effort to group villages so that policy making is right on target. One of the algorithms *clustering* that is commonly used is the *K-Means algorithm* because it is quite simple, easy to implement, and has the ability to group large data groups very quickly. However, the *K-Means algorithm* has a weakness in determining the center *cluster* initial given. Initialization of centers *cluster* randomly may result in formation *clusters* changing (inconsistent). For this reason, the *K-Means* method needs to be combined with the hierarchical method in determining the center *cluster* initial. This combination method is called *Hierarchical K-Means* which is a combination of methods *hierarchical* and *partitioning*, where the process is *hierarchical* used to find the initial center initialization *cluster* and the process *partitioning* to get the *cluster* optimal. The hierarchical method used in this study is the method *single linkage*. Based on the method *Elbow*, the recommended amount of  $k$  for this study is  $k = 4$ . The combination of the *single linkage* and *k-means algorithms* with  $k = 4$  in this study results in a value *silhouette coefficient* of 0.685 which is a feasible or appropriate cluster category, while the evaluation measurement by *Davies The Bouldin Index* yielded a value of 0.577.

**Keywords:** *Cluster, Davies Bouldin Index, K-means, Silhouette coefficient, Single linkage, Prediction*

## Pendahuluan

Kabupaten Pemalang merupakan kabupaten yang luas wilayahnya mencapai 1.115,30 km<sup>2</sup>. Berdasarkan letak administratif, Kabupaten Pemalang terdiri atas 14 kecamatan, 211 desa dan 11 kelurahan[1]. Setiap desa memiliki kondisi ekonomi, ciri dan tipologi lingkungan, akses fasilitas umum yang berbeda dan akan mengalami perubahan seiring dengan kemajuan pembangunan suatu desa. Pemanfaatan potensi yang ada pada setiap daerah perlu diperhatikan, karena melalui pengoptimalan potensi desa dapat mengangkat angka kemiskinan di suatu daerah[2].

Kabupaten Pemalang menduduki tingkat ke-4 teratas dari 35 Kabupaten/kota yang berada di Provinsi Jawa Tengah yang memiliki presentase kemiskinan tertinggi atau dibawah rata – rata sehingga menurut kriteria tipologi daerah termasuk dalam daerah relative tertinggal[1]. Berdasarkan data yang diperoleh dari Badan Pusat Statistik Kabupaten Pemalang tahun 2019, jumlah penduduk miskin di Kabupaten Pemalang sebanyak 200.670 jiwa dengan persentase sebesar 15,41 persen. Kabupaten Pemalang menempati posisi ke empat teratas dalam persentase kemiskinan di Provinsi Jawa Tengah.



Gambar 1.1 Presentase Kemiskinan di Provinsi Jawa Tengah (2019)

Berdasarkan data dari Badan Pusat Statistik menunjukkan kemiskinan sebagian besar dialami oleh masyarakat pedesaan selain itu rendahnya tingkat pelayanan infrastruktur di kawasan pedesaan hal inilah yang menjadi latar belakang kebijakan program pembangunan yang akan dilaksanakan[3]. Permasalahan umum yang dihadapi dalam kehidupan di setiap desa tentu berbeda – beda. Upaya mencapai tujuan pemerintah untuk menciptakan desa mandiri beberapa hal yang sangat perlu untuk dilakukan dan dikembangkan adalah mengutamakan pembangunan sarana dan prasarana yang ada di desa[3]. Pemerataan pembangunan perlu dilakukan untuk memperbaiki masalah kemiskinan[4]. Oleh karena itu, diperlukan sebuah upaya pengelompokan desa agar pengambilan kebijakan menjadi tepat sasaran[5]. Pemerintah pusat mulai memprioritaskan pembangunan dan pemberdayaan masyarakat di suatu daerah sampai tingkat desa agar tidak semakin tertinggal[6].

Dalam pelaksanaannya diperlukan identifikasi kondisi social masyarakat di setiap wilayah agar pengambilan keputusan atau kebijakan mengenai pembangunan bias tepat sasaran[7]. Ketepatan dalam pengidentifikasian target *group* atau target area menjadi prasyarat dalam keberhasilan program-program pembangunan[8]. Perencanaan pembangunan tidak terlepas dari kondisi masyarakat dan potensi daerah yang dimiliki. Adanya kebijakan pembangunan yang tepat sasaran dapat memberikan dampak positif terhadap jalannya roda perekonomian di suatu wilayah.

Berdasarkan latar belakang tersebut, akan dilakukan penelitian mengenai analisis *cluster* untuk pengelompokan desa di daerah Kabupaten Pemalang berdasarkan data Potensi Desa Kabupaten Pemalang. Analisis *cluster* adalah salah satu metode yang bertujuan untuk mengelompokkan data berdasarkan kemiripan karakteristik yang dimiliki dalam satu *cluster* yang sama. Potensi desa diamati berdasarkan ketersediaan infrastruktur serta keadaan sosial yang dimiliki oleh desa/kelurahan/nigari.

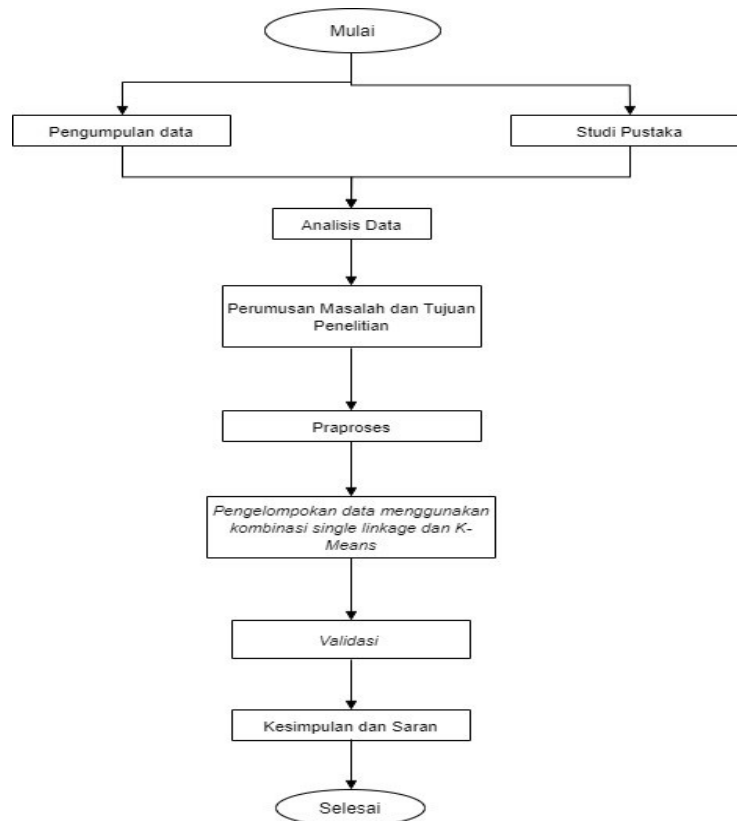
Salah satu algoritma *clustering* yang banyak digunakan adalah algoritma *K-Means* karena cukup sederhana, mudah diimplementasikan, dan memiliki kemampuan untuk mengelompokkan data yang besar dengan sangat cepat[9]. Namun algoritma *K-Means* memiliki kelemahan pada penentuan pusat *cluster* awal yang diberikan[10]. Pada *K-Means*, pusat *cluster* awal ditentukan secara acak atau random. Inisialisasi pusat *cluster* secara acak dapat memberikan hasil pembentukan *cluster* yang berubah-ubah (tidak konsisten) sehingga terkadang membutuhkan waktu yang cukup lama dalam melakukan iterasi untuk menemukan output *cluster* akhir. Inisialisasi pusat *cluster* yang tidak konsisten juga dapat menyebabkan masalah *local optimum* pada proses *K-Means* konvensional[11]. Untuk itu, metode *K-Means* perlu dikombinasikan

dengan metode hirarki dalam penentuan pusat *cluster* awal[10]. Metode kombinasi ini disebut sebagai *Hierarchical K-Means* yang merupakan penggabungan antara metode *hierarchical* dan *partitioning*, di mana proses *hierarchical* digunakan untuk mencari inisialisasi pusat *cluster* awal dan proses *partitioning* untuk mendapatkan *cluster* yang optimal. Metode hirarki yang akan diimplementasikan pada penelitian ini adalah metode *single linkage*.

Penelitian analisis cluster telah dilakukan oleh Alfini dkk., 2012 yang melakukan perbandingan antara metode *Hierarchical Clustering*, *K-Means*, dan gabungan keduanya dalam pengelompokan data problem kerja praktek jurusan Teknik Industri ITS. Berdasarkan penelitian yang dilakukan, diperoleh hasil bahwa kombinasi algoritma *hierarchical clustering* dan *k-means* menghasilkan pengelompokan data yang lebih baik dibandingkan *k-means* biasa. Dari kombinasi tersebut, diketahui kombinasi *single linkage* dan *k-means* yang menghasilkan *cluster* terbaik dibanding metode hirarki lainnya dilihat dari penjumlahan nilai *silhouette coefficient* terbesar yang dihasilkan[10]. Berdasarkan penelitian yang telah dilakukan sebelumnya, maka penggunaan algoritma *K-Means* memerlukan kombinasi untuk penentuan pusat cluster awal serta menentukan jumlah *k* yang digunakan untuk mencapai hasil cluster yang lebih baik. Pada penelitian ini melakukan *clustering* data menggunakan kombinasi *single linkage* dan *k-means*, penentuan jumlah *k* yang digunakan diperoleh dari metode *elbow* kemudian hasil *cluster* divalidasi menggunakan *silhouette coefficient* dan *Davies Boullidin Index*.

### Metode Penelitian

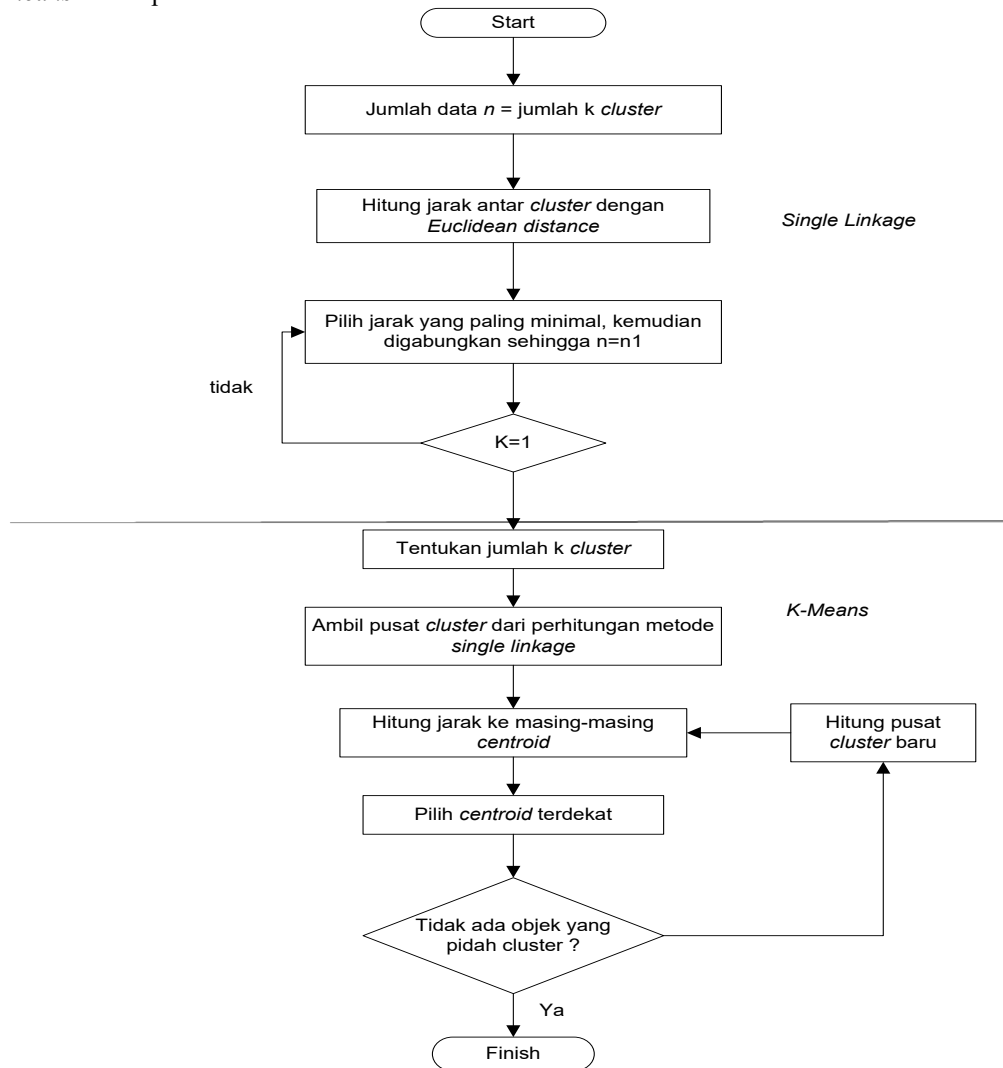
Berikut adalah flowchart penelitian yang berisi tahapan – tahapan penelitian yang dilakukan oleh peneliti :



**Gambar 1** Tahapan – Tahapan Penelitian

1. Pengumpulan data dan studi Pustaka, teknik pengumpulan data yang digunakan data dilakukan secara sekunder, selain itu dilakukan wawancara dengan pihak Badan Pusat Statistik Kabupaten Pemalang untuk menghimpun informasi mengenai potensi desa dengan parameter yang terlibat didalamnya dan studi pustaka dilakukan untuk menggali informasi lebih lanjut mengenai hal yang dibutuhkan saat penelitian.

2. Perumusan masalah dan tujuan penelitian, perumusan masalah serta tujuan penelitian yang ingin dicapai dalam penelitian ini yang didasarkan pada hasil analisa data pada tahap sebelumnya.
3. Praproses bertujuan untuk mempermudah dalam pemahaman terhadap isi record. Praproses yang dimaksud yaitu menyamakan satuan data atau menstandarisasikan data, untuk mengetahui adanya multikolinearitas antar variable dengan uji multikolinearitas menghitung nilai VIF (*Variance Inflation Factors*), serta mengukur kecukupan sampling secara menyeluruh dengan uji *Kaiser Mayer Olkin* (KMO) .
4. Penerapan metode, pada penerapan metode ini akan diuraikan langkah-langkah sistematis dan terarah menggunakan urutan metode *single linkage* yang di gunakan di tahap awal sedangkan *k-means* di tahap akhir.



**Gambar 2** Kombinasi *Single Linkage* dan *K-Means*

5. Validasi, pada proses validasi dilakukan pengukuran kualitas cluster yang terbentuk menggunakan *silhouette coefficient* dan *Davies Boullidin Index*.
6. Kesimpulan dan saran, penarikan kesimpulan dan saran berdasarkan hasil penelitian yang telah dilakukan

### *Asumsi Analisis Cluster*

Asumsi analisis cluster diperlukan untuk memeriksa variable – variable yang ada dan memastikan apakah variable layak digunakan dalam penelitian atau tidak. Dalam hal ini untuk mendeteksi multikolinearitas atau korelasi antar variable predictor menggunakan VIF (*Variance Inflation Factors*) dan untuk mengecek sampel yang mewakili atau sampel representative menggunakan uji Kaiser Mayer Olkin (KMO).

#### a. Multikolinearitas

Multikolinieritas merupakan suatu keadaan dimana terdapat korelasi antar variabel prediktor [12]. Salah satu cara yang dapat digunakan untuk mendeteksi adanya multikolinieritas yaitu dengan menghitung nilai VIF (*Variance Inflation Factors*) pada variabel yang diteliti [12] VIF dapat dihitung dengan rumus :

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

dengan  $R^2$  yang merupakan koefisien determinasi variabel dependen dengan variabel independen [13]. Jika nilai VIF lebih besar dari 10 menunjukkan adanya multikolinieritas antara variabel-variabel prediktor.

#### b. Sampel Yang Mewakili

Sampel yang mewakili atau sampel representatif merupakan sampel yang merepresentasikan atau mewakili populasi yang ada. Untuk mendeteksi hal tersebut dapat dilakukan dengan uji Kaiser Mayer Olkin (KMO). Sampel dapat dikatakan mewakili populasi atau sampel representative jika nilai KMO berada direntang 0,5 sampai dengan 1.

### Penentuan Jumlah Cluster Optimum

Pada metode partitioning, jumlah atau banyaknya cluster ( $k$ ) ditentukan terlebih dahulu. Penentuan jumlah cluster menggunakan metode Elbow [14]. Metode *Elbow* adalah metode yang digunakan untuk memperoleh informasi mengenai jumlah cluster terbaik yang sebaiknya digunakan pada penelitian dengan cara melihat presentase hasil berbanding antara jumlah cluster yang akan membentuk siku pada suatu titik [15]. Penentuan jumlah cluster pada metode ini dapat dihasilkan dari perbandingan hasil *Sum of Square Error* (SSE) pada masing - masing jumlah cluster dengan rumus SSE sebagai berikut (Irwanto dkk., 2012, seperti dikutip Merliana, 2015:17):

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

dengan,

$x_i$  = data pada objek ke- $i$

$c_k$  = *centroid cluster*  $C_k$

Hasil *Sum of Square Error* (SSE) akan di representasikan dengan grafik sebagai sumber informasi [15]

### Pemilihan Ukuran Jarak

Ukuran jarak digunakan untuk mengukur kemiripan antara dua objek. Salah satu ukuran jarak yang umum digunakan dalam *clustering* adalah jarak *Euclidean* [16]. Jarak *Euclidean* merupakan jarak dari suatu garis lurus yang menghubungkan antar objek [16]. Jarak *Euclidean* merupakan akar dari jumlah kuadrat perbedaan di dalam nilai untuk setiap variabel [12] (Supranto, 2004) :

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ik} - x_{jk})^2} \quad (3)$$

dengan :

$d(i, j)$  = Jarak dari data  $i$  ke data  $j$

$x_{ik}$  = Data  $i$  pada atribut data  $k$

$x_{jk}$  = Data ke- $j$  pada atribut data  $k$

### Metode Single Linkage

*Single linkage* adalah proses *clustering* yang didasarkan pada jarak terdekat antar objeknya (*minimum distance*). Metode ini akan menggabungkan dua buah objek yang memiliki jarak terdekat menjadi satu kelompok [12]. Anggota suatu cluster akan bertambah jika ada suatu variable yang memiliki jarak terdekat pada suatu cluster sehingga akan terjadi proses pengelompokkan secara hirarki [16]. Hasil dari metode *single linkage* dapat disajikan dalam bentuk diagram pohon atau dendogram.

Untuk menentukan jarak antar cluster menggunakan metode *single linkage* dapat dilakukan dengan perhitungan berikut :

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \quad (4)$$

di mana  $d_{UW}$  dan  $d_{VW}$  merupakan jarak *cluster* U dengan *cluster* W dan jarak *cluster* V dengan *cluster* W[12].

### Metode K-Means

Algoritma *K-Means* adalah salah satu algoritma *clustering* yang digunakan untuk mengelompokkan data berdasarkan titik pusat *cluster* (*centroid*) yang dekat dengan data, yang bertujuan untuk memaksimalkan kemiripan data dalam satu *cluster* serta meminimalkan kemiripan data antar *cluster*[17]. Menurut [18], untuk melakukan *clustering* dengan metode *K-Means* adalah sebagai berikut :

1. Menentukan jumlah *cluster* data atau k yang akan digunakan
2. Menentukan titik *centroid* atau pusat *cluster*.
3. Alokasikan semua data atau objek ke dalam *cluster* yang jaraknya paling dekat.
4. Hitung ulang pusat *cluster* dengan keanggotaan yang sekarang.

Tugaskan lagi setiap objek memakai *cluster* yang baru. Apabila pusat *cluster* yang diperoleh tidak berubah maka akan proses selesai. Jika tidak maka akan kembali ke langkah 3.

### Gabungan Metode Single Linkage Dengan K-Means 28

Metode *single linkage* merupakan salah satu dari metode *Hierarchical Clustering*, proses *hierarchical* digunakan untuk menentukan titik pusat *cluster* awal yang selanjutnya titik pusat *cluster* tersebut digunakan untuk proses pengelompokan data menggunakan metode *k-means*[10] [19] Gabungan metode ini dikembangkan dengan tujuan meningkatkan kualitas pengelompokan data, waktu yang dibutuhkan untuk proses pengelompokan lebih cepat terutama untuk big data.[20]

### Validasi

Setelah pengelompokan data maka dilakukan evaluasi hasil pengelompokan menggunakan validasi *cluster*. Validasi *cluster* dilakukan untuk mengukur kualitas *cluster* yang dihasilkan dari pengelompokan yang telah dilakukan[21]. Penelitian ini menggunakan validasi adalah *Silhouette Coefficient* dan *Davies Boulldin Index*.

*Silhouette Coefficient* adalah salah satu metode validasi *cluster* yang digunakan untuk melihat tingkat kualitas dan kekuatan *cluster* suatu objek di dalam suatu *cluster*[22]. Nilai *Silhouette Coefficient* dirumuskan sebagai berikut :

$$s(i) = \frac{(b(i) - a(i))}{\text{Max}\{a(i), b(i)\}} \quad (5)$$

dengan,

$a(i)$  = rata - rata jarak objek ke- $i$  terhadap semua objek di klaster A

$b(i)$  = nilai minimal rata - rata jarak objek ke- $i$  terhadap semua objek di klaster lain.

Tabel berikut menyajikan kriteria pengukuran nilai *silhouette coefficient* [21]

Tabel 1 Kriteria Pengukuran Nilai *Silhouette coefficient*

<i>Silhouette Coefficient</i>	Kriteria
0,71-1,00	<i>Cluster</i> yang kuat
0,51-0,70	<i>Cluster</i> yang layak atau sesuai
0,26-0,50	<i>Cluster</i> yang lemah
$\leq 0,25$	Tidak dapat dikatakan sebagai <i>Cluster</i>

sumber: Kaufmaan dan Rousseeuw,2005:85

*Davies Boulldin Index* (DBI) pertama kali diusulkan di tahun 1979 oleh David L. Davies dan Donald W. Boulldin . DBI ini mempunyai skema evaluasi *cluster* internal baik atau tidaknya *cluster* yang dihasilkan dapat dilihat dari kuantitas dan kedekatan antar data dari hasil *cluster*. [23]

DBI dilakukan dengan cara yaitu memaksimalkan jarak didalam sebuah *cluster* serta meminimalkan jarak diluar *cluster*[24]. Semakin kecil nilai DBI menunjukkan *cluster* yang dihasilkan optimal[25]. Nilai DBI dirumuskan pada persamaan :

$$DBI = \frac{1}{K} \sum_{i=1}^k \{R_i\} \quad (6)$$

dengan:

k: jumlah cluster

R<sub>i</sub>: Ukuran kemiripan

## Hasil dan Pembahasan

### Hasil Uji Multikolinearitas

Untuk pengujian multikolinieritas antar variabel dapat ditentukan dengan cara melihat nilai tolerance dan VIF yang dihasilkan. Berdasarkan tabel dibawah nilai tolerance lebih besar dari 0,10 serta nilai VIF < 10 oleh karena itu dapat disimpulkan bahwa tidak terjadi gejala multikolinearitas antar variabel.

Tabel 2 . Nilai VIF

#### Coefficients<sup>a</sup>

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	3.563	.000		755583097.871	.000		
Zscore(X1)	2.532	.000	1.000	215268258.029	.000	.161	6.192
Zscore(X2)	-9.143E-15	.000	.000	.000	1.000	.390	2.566
Zscore(X3)	-2.784E-15	.000	.000	.000	1.000	.583	1.715
Zscore(X4)	6.598E-15	.000	.000	.000	1.000	.517	1.936
Zscore(X5)	8.066E-15	.000	.000	.000	1.000	.479	2.088
Zscore(X6)	4.089E-15	.000	.000	.000	1.000	.622	1.609
Zscore(X7)	-4.732E-15	.000	.000	.000	1.000	.549	1.821
Zscore(X8)	-3.412E-16	.000	.000	.000	1.000	.142	7.046
Zscore(X9)	3.870E-15	.000	.000	.000	1.000	.108	9.251
Zscore(X10)	-1.585E-15	.000	.000	.000	1.000	.238	4.205
Zscore(X11)	2.055E-14	.000	.000	.000	1.000	.162	6.169
Zscore(X12)	2.654E-14	.000	.000	.000	1.000	.226	4.422
Zscore(X13)	-2.459E-15	.000	.000	.000	1.000	.745	1.343
Zscore(X14)	2.499E-15	.000	.000	.000	1.000	.924	1.082
Zscore(X15)	1.850E-15	.000	.000	.000	1.000	.842	1.187

### Hasil Uji Sampel Yang Mewakili Dengan KMO

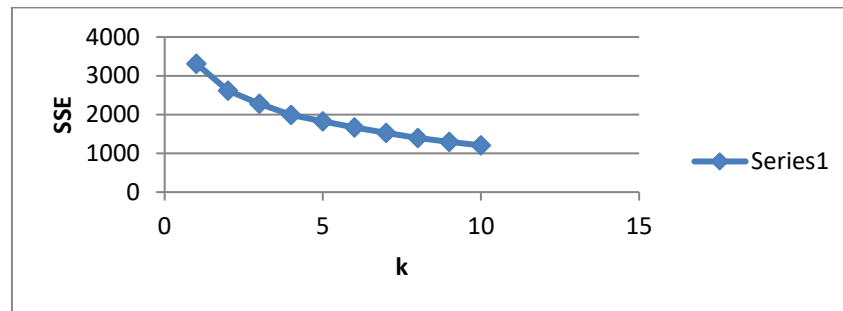
Berdasarkan Bartlett Test of Sphericity dengan Chi-Square 1940.466 (df 105) dengan nilai sig = 0,000 < 0,05 menunjukkan bahwa matriks korelasi bukan merupakan jenis matriks identitas oleh karena itu variable yang digunakan dapat dilakukan analisis komponen utama. Selain itu nilai KMO yang di hasilkan adalah 0,767 Berdasarkan nilai KMO dan *Bartlett's test* tersebut, dapat disimpulkan bahwa semua variabel layak untuk dianalisis dengan variabel yang dapat diolah lebih lanjut sebanyak 15 variabel.

Tabel 3 KMO and Bartlett's Test

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.767
Bartlett's Test of Sphericity	Approx. Chi-Square	1940.466
	df	105
	Sig.	.000

### Penentuan Jumlah Cluster Optimum Dengan Metode Elbow

Gambar dibawah merupakan plot metode elbow antara nilai SSE dengan banyaknya kluster (k) yang terbentuk. Nilai k atau banyaknya kluster yang akan dibentuk dapat dilihat ketika terjadi penurunan drastis dan terbentuk siku (elbow) untuk nilai SSE pada suatu nilai k. Pada plot di tersebut, nilai SSE mengalami penurunan drastis, dari k=1 sampai k=4 dan turun secara perlahan pada nilai k berikutnya hingga relatif stabil. Garis membentuk elbow pada k=4, dan setelahnya nilai SSE pada plot relatif stabil. Sehingga dapat diambil kesimpulan bahwa banyaknya kluster yang dibentuk adalah 4 kluster



Gambar 3 Grafik Metode Elbow

### Hasil Proses Single Linkage

Proses *clustering* menggunakan metode single linkage dimana setiap data akan dianggap sebagai cluster, jika terdapat data sejumlah n, kemudian k dianggap sebagai jumlah cluster, maka besarnya  $n = k$ . Selanjutnya akan dihitung jarak antar clusternya menggunakan Euclidean distance dari hasil perhitungan jarak tersebut dipilih jarak yang paling minimal untuk digabungkan menjadi sebuah *cluster* sehingga besarnya  $n = n - 1$  dan berhenti jika  $k = 1$ . Berdasarkan proses *single linkage* yang telah dilakukan, akan diambil titik pusat untuk  $k = 4$  dimana titik pusat ini akan digunakan untuk proses *clustering* selanjutnya menggunakan metode *k-means*.

cluster	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14
1	-0.071484	-0.040042	-0.035678	-0.079716	-0.068161	-0.029609	-0.058058	-0.059107	-0.093638	-0.079545	-0.061525	-0.048655	-0.020765	-0.067126
2	4.659472	1.368762	0.667762	5.879213	4.852905	2.531987	3.346474	0.345225	2.856152	5.012108	3.831438	3.221783	0.621407	0.618410
3	-0.222899	-0.476573	-0.799992	-0.187444	-0.513886	-0.452797	-0.551184	-0.228106	-0.163209	-0.187369	-0.206372	-0.868227	0.581306	12.875234
4	1.756441	5.059434	6.538780	-0.151883	1.346159	-0.718111	3.110253	12.016696	11.914233	2.412370	2.083068	1.761052	2.080420	-0.164027

Gambar 4 Hasil Single Linkage

### Hasil Gabungan Metode Single Linkage dan K-Means

Pada tahap ini titik pusat yang akan digunakan pada proses *clustering* dengan *k-means* berasal dari proses *single linkage*. Hasil dari proses *clustering* menggunakan metode gabungan antara *single linkage* dengan *k-means* ini yaitu sebagai berikut :

Klaster 0

Plakaran, Mandiraja, Walangsanga, Sima, Moga, Wangkelang, Kebanggan, Pepedan, Gendowang, Pakembaran, Warunpring, Karangdawa, Datar, Cibuyur, Mereng, Clekatakan, Baturisari, Penakir, Gunungsari, Jurangmangu, Gambuhan, Karangisari, Nyalembeng, Pulosari, Pagenteran, Siremeng, Cikendung, Gombong, Belik, Gunungtiga, Kuta, Badak, Gunungjaya, Simpurn, Mendelem, Beluk, Bulakan, Sikasur, Kalisaleh, Tundagan, Tlagasana, Bongas, Cikadu, Cawet, Medayu, Pagelaran, Bodas, Jojogan, Majalangu, Tambi, Watukumpul, Gapura, Majakarta, Wisnu, Longkeyang, Jatingarang, Gunungbatu, Pasir, Kwasen, Jatiroyom, Parunggalih, Payung, Cangak, Kemandungan, Kesesirejo, Babakan, Karangbrai, Jraganan, Kemandaran, Bodeh, Muncang, Kelangdepok, Pendowo, Sumurkidang, Wanarata, Pedagung, Suru, Banjarsari, Pegiringan, Karanganyar, Purana, Pabuaran, Sarwodadi, Bantarbolang, Sambeng, Glandang, Kuta, Kebon Gede, Paguyangan, Lenggerong, Kecepit, Gembyang, Mejangong, Penusupan, Banjaranyar, Karangmoncol, Semingkir, Semaya, Tanahbaya, Lodaya, Rembul, Krejo, Kalimas, Mangli, Kalitorong, Kejene, Gongseng, Banjarmasin, Surajaya, Pagongsoran, Sungapan, Paduraksa, Kramat,



Wanamulya, Mengori, Sewaka, Saradan, Bojongbata, Bojongnangka, Tambakrejo, Lawangrejo, Widuri, Danasari, Penggarit, Pener, Gondang, Jrakah, Sokawangi, Kejambon, Jebed Selatan, Jebed Utara, Cibelok, Kaligelang, Banjardawa, Banjaran, Sitemu, Pedurangan, Taman, Wanarejan Selatan, Wanarejan Utara, Beji, Kabunan, Asemtoyong, Kedungbanjar, Kendalsari, Widodaren, Karangasem, Sirangkang, Iser, Serang, Petarukan, Kalirandu, Pesucen, Panjunan, Temuireng, Pegundan, Bulu, Tegalmlati, Loning, Nyamplung Sari, Klareyan, Kendaldoyong, Kendalrejo, Sokawati, Tegalsari Barat, Tegalsari Timur, Kemuning, Karangtalok, Wonogiri, Blimbing, Ampelgading, Cibiyuk, Karangtengah, Banglarangan, Losari, Ujunggede, Jatirejo, Kebagusan, Sidokare, Tumbal, Pecangkalan, Sikayu, Purwosari, Purwoharjo, Kauman, Sidorejo, Lowa, Ambokulon, Gedeg, Gintung, Gandu, Sarwodadi, Susukan, Klegen, Wonokromo, Kebojongan, Kandang, Sukorejo, Botekan, Rowosari, Ambowetan, Pagergunung, Wiyorowetan, Samong, Tasikrejo, Bumirejo, Kaliprau, Kertosari, Pamutih, Padek, Blendung, Ketapang, Limbangan, Mojo, Pesantren

Klaster 1

Randudongkal, Kebondalem, Mulyoharjo, Pelutan, Sugiharwas

Klaster 2

Petanjungan

Klaster 3

Banyumudal

### **Perbandingan kualitas cluster menggunakan metode K-Means dan kombinasi Single Linkage dengan K-means**

Berikut adalah hasil dari penerapan algoritma *K-Means* dan Kombinasi algoritma *K-Means* dengan *Single Linkage* yang telah dilakukan menghasilkan informasi validasi *cluster* dari data potensi desa menggunakan variasi nilai *k*.

**Tabel 4** Hasil Perbandingan Evaluasi Model

<b>Jumlah Cluster (k)</b>	<b>K-means</b>		<b>Kombinasi Single Linkage &amp; K-Means</b>	
	<b>silhouette coefficient</b>	<b>Davies Boullidin Index.</b>	<b>silhouette coefficient</b>	<b>Davies Boullidin Index.</b>
3	0.346	1.206	0.670	0.727
<b>4</b>	<b>0.370</b>	<b>1.313</b>	<b>0.685</b>	<b>0.577</b>
5	0.367	1.012	0.299	0.986

Berdasarkan tabel diatas kualitas *cluster* yang dihasilkan oleh gabungan metode *single linkage* dan *k-means* lebih baik dibandingkan dengan *cluster* yang hanya dibentuk oleh metode *k-means*. Berdasarkan hasil evaluasi tersebut dapat disimpulkan bahwa kombinasi algoritma *single linkage* dan *k-means* dengan *k* = 4 menghasilkan nilai evaluasi yang paling baik dibandingkan dengan menggunakan algoritma *K-Means* tanpa kombinasi. Kombinasi antara metode *single linkage* yang digunakan untuk menentukan pusat *cluster* dan *k-means* terbukti dapat memperbaiki kualitas klaster yang dihasilkan.

### **Pengaruh Single Linkage dalam proses clustering**

Pada penelitian ini *single linkage* digunakan untuk mencari centroid yang akan digunakan pada proses *clustering k-means*. Dari hasil penelitian yang sudah dilakukan diperoleh hasil bahwa kualitas kluster yang dihasilkan dari penggabungan metode *single linkage* dan *k-means* rata – rata menghasilkan *cluster* yang layak sedangkan jika menggunakan metode *k-means* saja atau tanpa kombinasi semua *cluster* yang dihasilkan merupakan *cluster* yang lemah.

### **Kebijakan pembangunan yang dapat dilakukan**

Berdasarkan penelitian yang telah dilakukan untuk desa yang berada pada cluster nol Contoh pembangunan sarana prasarana di bidang ekonomi yang dapat dilakukan yaitu pembangunan minimarket atau pasar sebagai tempat bertransaksi untuk mendukung berkembangnya roda ekonomi ditingkat masyarakat, selain itu pembangunan fasilitas kesehatan dan distribusi tenaga kesehata didesa ini perlu dilakukan pemerataan agar semua masyarakat bias mendapatkan kemudahan dalam hal pengobatan.

Untuk desa yang berada pada *cluster* dua pembangunan infrastruktur yang dapat dilakukan yaitu membangun infrastruktur kesehatan contohnya puskesmas, jumlah tenaga kesehatan pada *cluster* ini tergolong minim sehingga diperlukan pemerataan tenaga kesehatan. Pembannan minimarket maupun pasar tradisionalpun perlu dilakukan untuk mendukung berkembangnya roda ekonomi masyarakat. Untuk *cluster* satu hanya perlu dilakukan pemeliharaan fasilitas desa karena desa yang berada pada *cluster* tersebut sarana prasarananya sangat baik sedangkan untuk desa yang berada pada *cluster* tiga pembangunan yang perlu dilakukan yaitu yaitu sector pariwisatanya, pembangunan sarana prasarana untuk mengembangkan potensi alam yang ada perlu dilakukan contohnya dengan membangun penginapan, membangun took atau warung disekitar tempat wisata, maupun akses menuju tempat wisata tersebut

### Kesimpulan

Berdasarkan penelitian yang dilakukan diperoleh bahwa *cluster* terbaik berdasarkan metode elbow yaitu  $k = 4$ . Hasil klaster dari kombnasi metode Single Linkage dan K-Means dapat diketahui bahwa *cluster* 0 merupakan kelompok desa yang kurang maju, ditinjau dari sarana prasarana desa, keadaan social penduduknya yang dicerminkan oleh rendahnya fasilitas pendidikan, kesehatan dibandingkan klaster yang lainnya. *cluster* 1 terdiri dari desa yang paling maju ditinjau dari faktor sarana dan prasarana desa serta faktor social ekonomi penduduknya. Umumnya desa – desa yang ada dalam kelompok ini merupakan kategori yang dekat dengan pusat kota. *Cluster* 2 merupakan kelompok desa yang memiliki ciri daerah yang paling tertinggal dibandingkan dengan kelompok lainnya. *Cluster* 3 adalah adalah kelompok desa yang terdiri dari desa yang cukup maju ditinjau dari faktor sarana dan prasarana desa, termasuk faktor social ekonomi penduduknya, namun secara keseluruhan nilai faktornya masih dibawah klaster 1.

Kualitas *cluster* yang dihasilkan oleh gabungan metode *single linkage* dan *k-means* lebih baik dibandingkan dengan *cluster* yang hanya dibentuk oleh metode *k-means*. Berdasarkan hasil evaluasi tersebut dapat disimpulkan bahwa kombinasi algoritma *single linkage* dan *k-means* dengan  $k = 4$  menghasilkan nilai evaluasi yang paling baik.

### Daftar Pustaka

- [1] Badan Pusat Statistik, “Statistik Potensi Desa Kabupaten Pematang 2019,” pp. 1–131, 2019.
- [2] A. Supandi, A. Saefuddin, and I. D. Sulvianti, “Two step Cluster Application to Classify Villages in Kabupaten Madiun Based on Village Potential Data,” *Xplore J. Stat.*, vol. 10, no. 1, pp. 12–26, 2020, doi: 10.29244/xplore.v10i1.272.
- [3] T. Alya and S. Sjaf, “ANALISIS RESPON MASYARAKAT TERHADAP PENGELOLAAN DANA DESA UNTUK PEMBANGUNAN PEDESAAN ( Desa Pesantren , Kecamatan Ulujami , Kabupaten Pematang ) Analysis of Community Response towards The Management of Village Fund for Rural,” *Sains Komun. dan Pengemb. Masy.*, vol. 2, no. 3, pp. 317–332, 2018.
- [4] N. Irabawati, S. Wahyuningsih, and R. R. Syoer, “Perbandingan Metode C-Means dan Fuzzy C-Means Dalam pengelompokkan Wilayah Desa/Kelurahan di Kabupaten Kutai Kartanegara,” vol. 7, pp. 1–13, 2014.
- [5] W. Alwi and M. Hasrul, “Analisis Klaster Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat,” *J. MSA ( Mat. dan Stat. serta Apl. )*, vol. 6, no. 1, p. 35, 2018, doi: 10.24252/msa.v6i1.4782.
- [6] Z. S. Badu, “Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Dana Desa,” *J. Inform.*, no. November, 2016,
- [7] Y. B. Alam and A. N. Ambarwati, “Analisis Cluster Pada Desa / Kelurahan Di Kabupaten Wonosobo Berdasarkan Data Potensi Desa Tahun 2015,” no. February, pp. 351–358, 2017.
- [8] Soemartini and E. Supartini, “Analisis K-Means Cluster Untuk Pengelompokan Kabupaten / Kota Di Jawabarot Berdasarkan Indikator Masyarakat,” *Prosiding*, no. Knmp Ii, pp. 144–154, 2017.
- [9] A. S. Rizal and R. F. Hakim, “Metode K-Means Cluster Dan Fuzzy C-Means Cluster (Studi Kasus: Indeks Pembangunan Manusia Di Kawasan Indonesia Timur Tahun 2012),” *Pros. Semin. Nas. Mat. dan Pendidik. Mat. UMS 2015*, pp. 643–657, 2015, [Online]. Available: <https://publikasiilmiah.ums.ac.id/xmlui/handle/11617/5803>.
- [10] T. Alfina and B. Santosa, “Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Keduanya dalam Membentuk Cluster Data (Studi Kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS),” *Anal. PerbandinganMetode Hierarchical Clust. K-means dan Gabungan Keduanya dalam Clust. Data*, vol. 1, no. 1, pp. 1–5, 2012, doi: 10.12962/j23373539.v1i1.1794.
- [11] S. Q. Ye Hanmin, Lv Hao, “An Improved Semi-Supervised K-Means Clustering Algorithm,” pp. 5–8.

- 
- [12] A. N. Fathia and R. Rahmawati, "Analisis klaster kecamatan di kabupaten semarang berdasarkan potensi desa menggunakan metode ward dan single linkage," vol. 5, pp. 801–810, 2016.
- [13] D. N. Gujarati and D. C. Porter, "Dasar-dasar ekonometrika," *Jakarta: Erlangga*, 2006.
- [14] S. H. Ghwanmeh, "Applying Clustering of Hierarchical K-means-like Algorithm on Arabic Language," vol. 1, no. 8, pp. 2390–2394, 2007.
- [15] N. Putu, E. Merliana, and A. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means," pp. 978–979.
- [16] K. U. Mengelompokkan and K. Di, "LINKAGE , COMPLETE LINKAGE , AVERAGE LINKAGE , DAN JAWA TENGAH BERDASARKAN PRODUKSI PALAWIJA," 2017.
- [17] R. A. Asroni, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Ilm. Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2015, doi: 10.1038/hdy.2009.180.
- [18] B. Santosa, "Data mining teknik pemanfaatan data untuk keperluan bisnis," *Yogyakarta Graha Ilmu*, vol. 978, no. 979, p. 756, 2007.
- [19] L. Rahmawati, S. Widya Sihwi, and E. Suryani, "Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus: Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret)," *J. Teknol. Inf. ITSmart*, vol. 3, no. 2, p. 66, 2016, doi: 10.20961/its.v3i2.654.
- [20] T. Xu, H. Chiang, G. Liu, and C. Tan, "Hierarchical K-means Method for Clustering Large-Scale Advanced Metering Infrastructure Data," vol. 8977, no. c, 2015, doi: 10.1109/TPWRD.2015.2479941.
- [21] D. F. Azuri, Zulhanif, and R. S. Pontoh, "Pengelompokan Kabupaten/Kota Di Pulau Jawa Berdasarkan Pembangunan Manusia Berbasis Gender Menggunakan Bisecting K-Means," *Peran Penelit. Ilmu Dasar dalam Menunjang Pembang. Berkelanjutan*, pp. 78–83, 2016, [Online]. Available: [http://riset.fmipa.unpad.ac.id/proc\\_semnamipa2016/017.078-083DilaFitriani.pdf](http://riset.fmipa.unpad.ac.id/proc_semnamipa2016/017.078-083DilaFitriani.pdf).
- [22] D. A. Wicaksana, P. P. Adikara, and S. Adinugroho, "Clustering Dokumen Skripsi Dengan Menggunakan Hierarchical Agglomerative Clustering," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 12, 2018.
- [23] D. Jollyta, S. Efendi, M. Zarlis, and H. Mawengkang, "Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 918, 2019, doi: 10.30645/senaris.v1i0.100.
- [24] M. I. Komputer, F. I. Komputer, and U. D. Nuswantoro, "Algoritma Cluster Dinamik Untuk Optimasi Cluster Pada Algoritma K-Means Dalam Pemetaan Nasabah Potensial," *J. Intell. Syst.*, vol. 1, no. 1, pp. 33–36, 2015.
- [25] F. Ramdhani and A. Hoyyi, "Pengelompokan Provinsi Di Indonesia Berdasarkan Karakteristik Kesejahteraan Rakyat Menggunakan Metode K-Means Cluster," *J. Gaussian*, vol. 4, no. 4, pp. 875–884, 2015.