

# Pengelompokan Desa Menggunakan K-Means untuk Penyelenggaraan Penanggulangan Bencana Banjir

Shelladita Fitriyani Susilo<sup>1</sup>, Asep Jamaludin<sup>2</sup>, Intan Purnamasari<sup>3</sup>

Universitas Singaperbangsa Karawang

e-mail: <sup>1</sup>shelladita.16207@student.unsika.ac.id, <sup>2</sup>asep.jamaludin@staff.unsika.ac.id,

<sup>3</sup>intan.purnamasari@staff.unsika.ac.id

Diterima: 25 Juni 2020; Direvisi: 08 September 2020; Disetujui: 25 September 2020

## Abstrak

Penyelenggaraan penanggulangan bencana merupakan suatu agenda kewajiban lembaga yang dinamakan BPBD. Perencanaan yang tidak terstruktur menyebabkan penentuan komponen didalamnya menjadi kurang optimal, seperti tempat, sumber daya manusia, transportasi. Penentuan tempat menjadi poin utama yang dapat mempengaruhi komponen lain dalam perencanaan. Ketersediaan data mengenai tempat berupa rekap desa rawan bencana dimiliki oleh bidang I Kesiapsiagaan Bencana. Dari rekap data kejadian dan profil desa dilakukan pengelompokan yang bertujuan untuk menemukan karakteristik yang dapat membantu dalam penentuan tempat kegiatan. Metode K-Means dipilih untuk memetakan desa sesuai karakteristik dengan jumlah cluster sesuai jumlah optimal yang lebih baik antara metode elbow dan silhouette. Dihasilkan clustering optimal dengan jumlah cluster 7 hasil penentuan nilai k dari metode elbow. Dengan ukuran pengelompokan Cluster 1 = 1 desa, Cluster 2 = 2 desa, Cluster 3 = 19 desa, Cluster 4 = 4 desa, Cluster 5 = 5 desa, Cluster 6 = 1 desa, Cluster 7 = 1 desa.

**Kata kunci:** Clustering, Elbow, K-Means, Silhouette

## Abstract

Disaster management is an institutional obligation agenda called BPBD. Unstructured planning causes the determination of components in it to be less than optimal, such as place, human resources, transportation. Determination of place is the main point that can affect other components in planning. The availability of data on places in the form of recapitalization of disaster-prone villages is owned by the Disaster Preparedness I field. The recapitulation of village events and profiles is grouped to find characteristics that can help determine the location of activities. The K-Means method is used to map villages according to characteristics with the number of clusters according to the optimal number that is better between the elbow and silhouette methods. Optimal clustering is produced with the number of clusters 7 results of determining the value of k from the elbow method. The grouping size Cluster 1 = 1 village, Cluster 2 = 2 villages, Cluster 3 = 19 villages, Cluster 4 = 4 villages, Cluster 5 = 5 villages, Cluster 6 = 1 village, Cluster 7 = 1 village.

**Keywords:** Clustering, Elbow, K-Means, Silhouette

## 1. PENDAHULUAN

Banjir di kabupaten Tegal merupakan bencana alam yang sering terjadi pada bulan awal tahun. Pada Januari 2020 telah tercatat laporan kejadian banjir yang melanda 21 desa di beberapa kecamatan. Berdasar pada UU Nomor 24 Tahun 2007 tentang Penanggulangan Bencana yang

---

menjelaskan banjir merupakan bencana yang ditanggulangi oleh BPBD. Dalam penanggulangan, BPBD kabupaten Tegal melakukan kegiatan sosialisasi, penyiapan daerah tanggap bencana, dan edukasi terkait pencegahan atau pemantauan langsung. Namun kegiatan seringkali diputuskan secara mendadak. Kurangnya persiapan dalam penentuan tempat dapat mengulur waktu, karena perhatian terhadap ketersediaan sumber daya manusia serta transportasi dilakukan secara mendadak pula. Pemilihan Kegiatan pun sering menjadi kendala karena acuan berpusat pada anggaran. Desa rawan bencana banjir menjadi prioritas BPBD dalam menentukan tempat sehingga digunakan data desa rawan bencana untuk memilah desa sesuai karakteristiknya. Dengan pemilahan desa tentu dapat menentukan kegiatan penyelenggaraan penanggulangan bencana banjir sesuai dengan data desa dan anggaran yang diberikan.

Penambahan data untuk mengelompokan atau memilah desa telah dilakukan sebelumnya dengan menghasilkan kelompok desa berdasar provinsi melalui jumlah pengguna dan sumber listrik [1]. Penambahan data yang digunakan untuk membagi data menjadi suatu kelompok yang homogen ialah *clustering* [2]. *K-Means* menjadi salah satu algoritme pengelompokan dalam menghasilkan rekomendasi pekerjaan dengan mengelompokkan data penduduk metode ini memiliki nilai akurasi yang dikategorikan cukup [3].

## 2. METODE PENELITIAN

### 2.1 Dataset

Data berhasil dikumpulkan pada tahapan pemahaman data setelah ditentukan latar belakang dan tujuan penelitian pada tahapan pemahaman bisnis. Data yang digunakan merupakan data sekunder karena melalui wawancara, studi pustaka dan jelajah internet, yaitu dari laporan harian kejadian bencana banjir, web resmi acuan BPBD seperti *bnpb.cloud* dan *prodeskel.binpemdes.kemendagri.go.id*. Data desa rawan bencana kabupaten Tegal yang dikumpulkan berjumlah 74 desa dengan 20 atribut dengan beberapa data yang dilihat pada Gambar 1.

KECAMATAN	DESA	jml_pwa	KK	kepadatan	sawah	bukan_sawah	banjir	jml_kejadian	min	maks	Kecelakaan			Tersak Bencana rak	Bekap barang		Droping air (rata- rata)	
											KEJADIAN SUNAM	KEJADIAN GEMPA BUNYI atau GEMPA BUNYI	KEJADIAN GEMPA BUNYI atau GEMPA BUNYI		min	maks		
1. Kraman	1. Patehaya	5.761	3953	304,22	13.153	30.00	30.00	0	0	0	0	0	0	0	0	0	0	0
	2. Banting	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	3. Kemiripati	4917	1288	16,87	61.326	26.99	26.99	0	0	0	0	0	0	0	0	0	0	0
	4. Banting	1440	4801	333,24	95.367	143.31	143.31	0	0	0	0	0	0	0	0	0	0	0
	5. Banting	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
2. Suralaga	1. Suralaga	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	2. Suralaga	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	3. Suralaga	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	4. Suralaga	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	5. Suralaga	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
3. Dukuhrejo	1. Dukuhrejo	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	2. Dukuhrejo	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	3. Dukuhrejo	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	4. Dukuhrejo	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	5. Dukuhrejo	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
4. Tilai	1. Tilai	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	2. Tilai	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	3. Tilai	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	4. Tilai	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	5. Tilai	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
5. Banting	1. Banting	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	2. Banting	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	3. Banting	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	4. Banting	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0
	5. Banting	3334	1811	179,70	25.124	111	111	0	0	0	0	0	0	0	0	0	0	0

Gambar 1. Rekap Laporan Kejadian Daerah rawan bencana

Berdasarkan acuan awal BPBD, data yang dipilih ialah 33 desa rawan bencana banjir dan 8 atribut yang sering digunakan sebagai keterangan saat terjadi bencana banjir. Antara lain KK, Kepadatan, Sawah, Bukan sawah, Banjir, jml\_kejadian, min, dan maks. Pembentukan *dataset* berkualitas dilanjutkan pada tahapan persiapan dengan melakukan imputasi untuk mengatasi *missing value*. Imputasi menggunakan *mean*, karena metode tersebut metode imputasi terbaik setelah 1-NN [4]. Menentukan nilai *mean* untuk imputasi pada data desa rawan banjir menggunakan formula sebagai berikut [5] :

$$Mean = \frac{\text{jumlah data}}{\text{banyak data}} \tag{1}$$

Tabel 1. Hasil imputasi menggunakan *Mean*.

Atribut	Jumlah data	Banyak data	Mean (Imputasi)
KK	6891	33	67
Kepadatan	978.03	33	30.564
Sawah	188404	33	5887.612
Bukan Sawah	40515	33	1266.104

Penggunaan persamaan (1) yaitu *mean* untuk penanganan *missing value* pada data desa Dampyak, sehingga menghasilkan imputasi seperti dalam Tabel 1. Penanganan selanjutnya yaitu *contract data* atau transformasi data. Penanganan tersebut dilakukan untuk mengubah data ke dalam bentuk lain sesuai kebutuhan. *Value* pada penelitian ini memiliki bentuk data yang besar sehingga perlu melakukan transformasi ke bentuk yang lebih kecil menggunakan transformasi nilai kedalam bentuk *z-score* [6]. Tabel 2 menjelaskan contoh proses transformasi pada salah satu *record* data pertama menggunakan persamaan (2).

$$z = \frac{x - \bar{x}}{s} \quad (2)$$

Dimana :

$x$  : data

$\bar{x}$  : rata-rata data

$s$  : standart deviasi

Tabel 2. Contoh transformasi data kedalam bentuk *z-score*.

Atribut	$x$	$\bar{x}$	$x - \bar{x}$	$s$	<i>z-score</i>
KK	300	210.85	89.152	359.71	0.247841592
Kepadatan	19.553	30.564	-11.011	25.861	-0.42576205
Sawah	161.99	5887.6	-5725.6	24627	-0.2325
Bukan Sawah	68.9	1266.1	-1197.2	5769.4	-0.207508675
Jml kejadian	3	1.6667	1.3333	1.0918	1.2212
min	50	66.364	-16.364	50.741	-0.3225
maks	110	102.42	7.5758	56.624	0.1338

## 2.2 Algoritme *K-Means*

Pada tahapan pemodelan dipilih teknik *clustering* dalam *data mining* guna tercapainya hasil sesuai perencanaan. Dalam setiap kelompok memiliki karakter yang sama namun antar kelompok diciptakan tingkat variasi yang kecil, sehingga memberikan karakteristik yang berbeda [7]. Pengelompokan menggunakan algoritme *K-Means*, sehingga berikut proses pengolahan *dataset* yang telah dipersiapkan sebelumnya :

1. Menentukan jumlah *cluster* ( $k$ )
2. Menentukan  $k$  yang menjadi titik pusat *cluster* (*centroid*) secara acak untuk awal perhitungan, selanjutnya akan ditentukan *centroid cluster* untuk iterasi berikutnya menggunakan persamaan (3).

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

Dengan:

$v$  : titik pusat kelompok

$x_i$  : objek data ke= $i$

$n$  : jumlah objek yang menjadi anggota *cluster*

- Menghitung jarak setiap objek data ke setiap *centroid* yang telah ditentukan menggunakan *Euclidean distance* [7] pada persamaan (4) . Tempatkan setiap objek sesuai dengan titik pusat *cluster* terdekat.

$$D(x_2, x_1) = \|x_2, x_1\|_2 = \sqrt{\sum_{j=1}^p (x_{2j} - x_{1j})^2} \tag{4}$$

Dengan :

$D$  : jarak kuadrat *Euclidean* antara objek ke  $x_2$  dan  $x_1$

$P$  : jumlah variabel cluster

$X_{2j}$  : data ke-2 pada variabel ke- $j$

$X_{1j}$  : data ke-1 pada variabel ke- $j$

- Melakukan iterasi untuk menentukan titik pusat *cluster* baru
- Ulangi Langkah ke-3 hingga titik pusat *cluster* pada titik yang sama

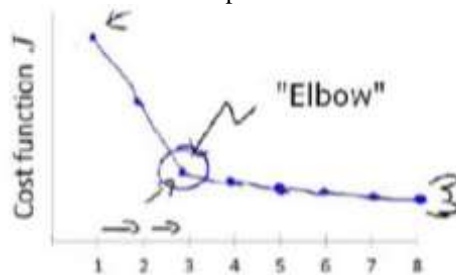
### 2.3 Metode *Elbow*

Tahapan pertama dalam algoritme *K-Means* yaitu menentukan nilai  $k$  [7]. Tahapan pemodelan dalam penelitian ini juga menentukan nilai  $k$  optimal berdasarkan pada hasil perbandingan penerapan metode *elbow* dan *silhouette*. Titik siku yang terbentuk dari hasil membandingkan nilai persen dari setiap jumlah cluster adalah cara metode *elbow* bekerja [7]. Perbandingan didapatkan dengan nilai *SSE* (*Sum of Square Error*) setiap kelompok. Penghitungan nilai *SSE* pada *K-Means* [8] dirumuskan pada persamaan (5).

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|X_i - C_k\|_2 \tag{5}$$

Dengan  $K$  merupakan banyak *cluster* yang dipilih,  $X_i$  merupakan jumlah objek, dan  $C_k$  merupakan jumlah anggota pada *cluster* ke- $k$ . Pada [7] Semakin kecil nilai *SSE* semakin besar nilai *cluster*, sehingga akan terbentuk siku saat terjadi penurunan drastic seperti pada Gambar 2.

<spasi>



Gambar 2. Grafik *Elbow*. Sumber : Eka Merliana et al., 2015 dalam [7]

### 2.4 Metode *Silhouette*

Cara menentukan *cluster* yang baik dengan metode ini yaitu dengan nilai rata-rata dari perhitungan *silhouette* yang tinggi [9]. Nilai *silhouette* ditentukan dengan perhitungan sebagai berikut [10]:

- Menghitung *average distance* objek ke- $o$  dengan semua objek dalam satu *cluster* menggunakan persamaan (6).

$$a(o) = \frac{\sum_{o' \in C_i, o \neq o'} dist(o, o')}{|C_i - 1|} \tag{6}$$

Dimana:

$o$  : Objek ke- $o$  pada *cluster*  $i$

$o'$  : Objek lain pada *cluster*  $i$  selain objek ke- $o$

$C_i$  : *Cluster*  $i$

$|C_i|$  : jumlah objek pada *cluster*  $i$

$dist(o, o')$  : jarak data  $o$  dengan data  $o'$

- Menghitung *average distance* objek ke- $o$  dengan semua objek dari *cluster* lain menggunakan persamaan (7) kemudian pilih dengan nilai terendah.

$$b(i) = \min_{C_j: 1 \leq j \leq k, j \neq i} \left\{ \frac{\sum_{o' \in C_j} \text{dist}(o, o')}{|C_j - 1|} \right\} \quad (7)$$

Dimana:

$o$  : Objek ke- $o$  pada *cluster*  $j$

$o'$  : Objek lain pada *cluster*  $j$  selain objek ke- $o$

$C_j$  : *Cluster*  $j$

$|C_j|$  : jumlah objek pada *cluster*  $j$

$\text{dist}(o, o')$  : jarak data  $o$  dengan data  $o'$

3. Pada objek  $k$  memiliki nilai *silhouette coefficient* dengan memasukkan nilai  $a(o)$  dan  $b(o)$  pada persamaan (8) :

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (8)$$

Dimana:

$S_i$  : Nilai *Silhouette coefficient* pada objek ke  $i$

$a_i$  : Nilai  $a(o)$  pada objek ke  $i$

$b_i$  : Nilai  $b(o)$  pada objek ke  $i$

### 2.5 Parameter *cluster* optimal

Penilaian ideal *cluster* jika  $V$  memiliki nilai yang minimum, yaitu dengan penghitungan  $V$  [6] pada persamaan (9). Ideal *cluster* dapat dilihat sesuai keterangan pada Tabel 3, yaitu *cluster* yang memiliki *variance within* minimum dan *variance between* maksimum.

$$V = \frac{\text{Variance within}}{\text{Variance between}} \times 100 \% \quad (9)$$

Tabel 3. Parameter penentuan hasil *clustering* k optimal

Parameter	Keterangan
<i>Variance withinss</i>	Semakin kecil <i>variance withinss</i> semakin baik pengelompokan
<i>Variance Beetweenss</i>	Semakin besar <i>variance betweenss</i> semakin baik pengelompokan

*Between sum of square* adalah Jarak antar *cluster* dan *Within Variance sum of square* adalah Jarak antara anggota dalam *cluster* [11]. Diperlukan nilai *variance within* yang minimum dan *variance between* maksimum untuk menentukan *cluster* ideal [6] yang bisa dihitung menggunakan persamaan (10), (11), dan (12). Berikut formula perhitungan *variance between* dan *variance within* menurut Keppel & Wickens (2004) dalam [12]:

$$\text{Variance} = \text{Mean Asuqres (MS)} = \frac{\text{sum of square}}{\text{derajat bebas (df)}} \quad (10)$$

$$\text{Variance between} = \frac{\text{sum of square between}}{\text{banyaknya kelompok} - 1} \quad (11)$$

$$\text{Variance within} = \frac{\text{sum of square within}}{\sum(\text{anggota per kelompok} - 1)} \quad (12)$$

## 3. HASIL DAN PEMBAHASAN

Sesuai dengan alur tahapan-tahapan yang dimiliki oleh CRISP-DM, penelitian ini menghasilkan kelompok desa dengan pembagian kelompok yang ideal. Data dilakukan pemilihan hingga menjadi *dataset* berkualitas. Kemudian diterapkan metode *K-Means* dengan memaksimalkan tahapan pemilihan k optimal. Perbandingan metode *elbow* dan *Silhouette* untuk

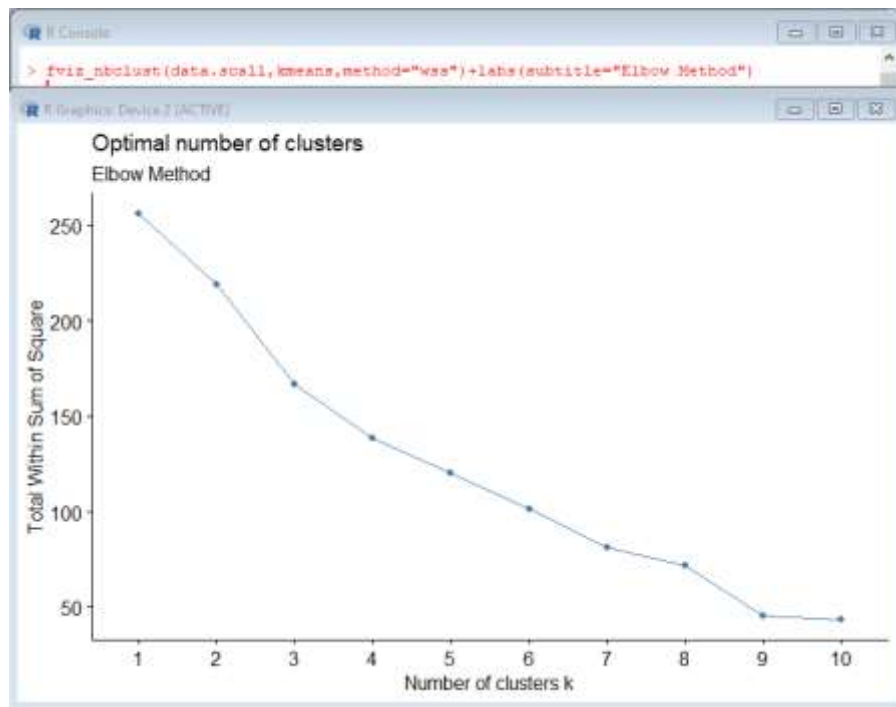
menentukan nilai k optimal. Pemodelan dilakukan dengan menggunakan *software* R. Penerapan pada R menghasilkan beberapa komponen dengan keterangan yang tertera pada Tabel 4. Indikator yang menjadi perbandingan hasil *clustering* ialah nilai *variance between* dan *variance withinss*.

Tabel 4. Hasil penerapan *K-Means* pada R sumber [13]

Komponen	Keterangan
<i>cluster</i>	Hasil analisis pengelompokan dari 1:k
<i>center</i>	Matriks titik pusat <i>cluster</i>
<i>totss</i>	Jumlah keseluruhan deviasi kuadrat
<i>Withinss</i>	Vektor dari <i>within-cluster sum of square</i> , suatu objek dalam <i>cluster</i>
<i>Tot.withinss</i>	Total <i>within-cluster sum of square</i>
<i>betweenss</i>	<i>Between-cluster sum of squares</i> , atau <i>totss-tot.withinss</i> .
<i>Size</i>	Jumlah anggota per <i>cluster</i>
<i>Iter</i>	Jumlah iterasi
<i>Ifault</i>	Jumlah masalah pada algoritma (untuk ahli)

### 3.1. Penentuan nilai k dengan metode *elbow*

Hasil nilai K optimal dengan metode *elbow* pada Gambar 3 ditentukan oleh garis yang mulai melandai setelah curam hingga membentuk siku pada titik k optimal [7]. Pada Gambar 3 garis mulai melandai setelah titik 7 sehingga titik k optimal menggunakan metode *elbow* ialah 7.



Gambar 3. Hasil nilai k metode *elbow*

Berdasarkan perhitungan *K-Means* pada R, perhitungan berhenti pada iterasi ke-4 dengan jumlah data per *cluster* yang di vektorkan pada Gambar 4 dengan komponen *iter* dan *size*. Iterasi tidak dilanjutkan karena algoritme telah mencapai kondisi dimana alokasi kembali pada titik data sebelumnya dan titik pusat *cluster* tetap sama [14]. Gambar 4 juga menunjukkan vektor *Within sum of square*, *Total Within sum of square* dan *Between sum of square* yang dihasilkan dengan nilai optimal 7 pada komponen *withinss*, *Tot.withinss*, dan *Betweenss*.

```

> resultm[4]
$withinss
[1] 0.00000 10.54243 30.05926 10.04201 16.19255 0.00000 0.00000

> resultm[5]
$tot.withinss
[1] 66.83625

> resultm[6]
$betweenss
[1] 189.1637

> resultm[7]
$size
[1] 1 2 19 4 5 1 1

> resultm[8]
$iter
[1] 4

```

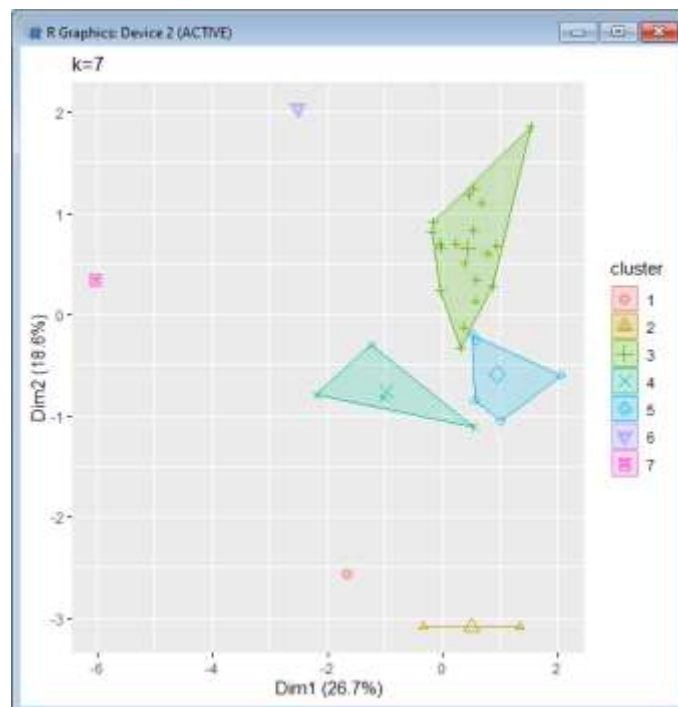
Gambar 4. Komponen hasil *K-Means* dengan metode *elbow*

Jumlah deviasi kuadrat (*sum of square*) yang dihasilkan dapat digunakan untuk menentukan varians (*variance*) yang dibagi dengan derajat bebas (*degrees of freedom* atau *df*) [12]. Nilai *variance* dari  $k=7$  yaitu sebagai berikut :

$$\text{Variance between} = \frac{189.1637}{7-1} = 27.0233857$$

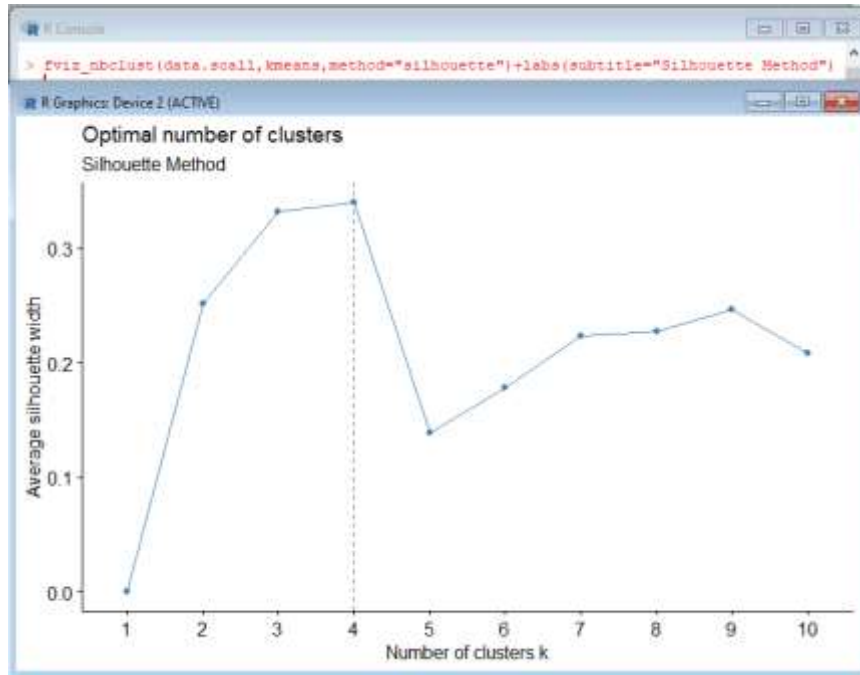
$$\text{Variance within} = \frac{66.83625}{((1-1)+(2-1)+(19-1)+(4-1)+(5-1)+(1-1)+(1-1))} = 2.570625$$

Penerapan *K-Means* menggunakan  $k=7$  yang menghasilkan *size* atau jumlah anggota per *cluster* berurutan yaitu 1, 2, 19, 4, 5, 1, 1 yang divisualisasikan pada Gambar 5



Gambar 5. Pemetaan menggunakan nilai  $k$  metode *elbow*

3.2. Penentuan nilai k dengan metode *silhouette*



Gambar 6. hasil nilai k metode *silhouette*

Sesuai dengan garis penunjuk yang memotong grafik pada Gambar 6, menunjukkan bahwa nilai k optimal penerapan metode *silhouette* ialah 4. Proses *clustering* dengan nilai k=4 menghasilkan *tot.withinss* (Total *withinss* sum of square) sebesar 136.384 dan *Betweenss* (*betweenss* sum of square) sebesar 119.616. Komponen hasil dari penerapan *K-Means* lainnya menggunakan k=4 dapat dilihat pada Gambar 7, seperti komponen *size* yang merupakan jumlah anggota dari 4 *cluster* secara berturut-turut, vektor *withinss* dan jumlah iterasi terakhir dalam mengalokasikan objek dengan perhitungan jarak.

```

> resultk[4]
$withinss
[1] 36.39376 62.76832 10.54243 26.67948

> resultk[5]
$tot.withinss
[1] 136.384

> resultk[6]
$betweenss
[1] 119.616

> resultk[7]
$size
[1] 6 23 2 2

> resultk[8]
$iter
[1] 3
    
```

Gambar 7. Komponen hasil *K-Means* dengan metode *silhouette*

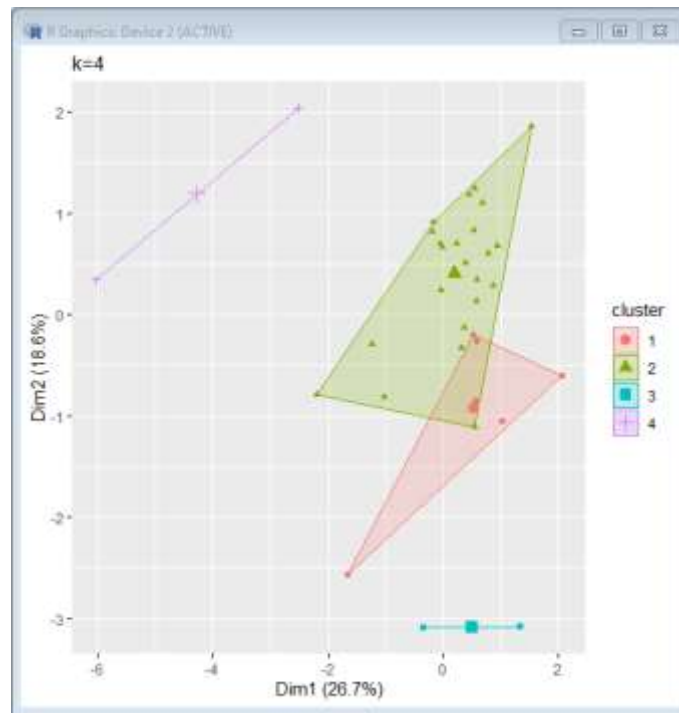
*Withinss* sum of square dan *between* sum of square yang dihasilkan kemudian dilakukan perhitungan untuk menentukan *variance* atau jarak objek [12]. *Variance between* dan *Variance within* yang dihasilkan dari 4 *cluster* menggunakan algoritme *K-Means* ialah sebagai berikut:



$$\text{Variance between} = \frac{119.616}{4-1} = 39.872$$

$$\text{Variance within} = \frac{136.384}{((6-1)+(23-1)+(2-1)+(2-1))} = 4.70289655$$

Size atau jumlah anggota per *cluster* berurutan hasil penerapan *K-Means* menggunakan nilai *k* 4 yaitu 1, 2, 19, 4, 5, 1, 1 divisualisasikan dalam bentuk pemetaan pada Gambar 8.



Gambar 8. Pemetaan menggunakan nilai *k* metode *silhouette*

### 3.3. Pembahasan

Pada Tabel 5 memperlihatkan rekap hasil perhitungan *variance* penerapan 2 metode penentuan nilai *k*, yaitu metode *elbow* dan *silhouette*.

Tabel 5. Perbandingan nilai *variance*

Metode	Cluster	Variance Within	Variance Between
<i>Elbow</i>	7	2.570625	27.0233857
<i>Silhouette</i>	4	4.70289655	39.872

Dari penentuan nilai *k* optimal dengan kedua metode dihasilkan *variance within* dan *variance between* dari metode *elbow* dan metode *silhouette* yang akan digunakan untuk mengevaluasi pengelompokan teroptimal, sebagai berikut :

Metode *Elbow*

$$V_{\text{Elbow}} = \frac{2.570625}{27.0233857} \times 100\% = 9.51259412 \sim 9.5\%$$

Metode *Silhouette*

$$V_{\text{Silhouette}} = \frac{4.70289655}{39.872} \times 100\% = 11.79498532 \sim 11.8\%$$

Dari perhitungan *V* masing-masing metode dihasilkan nilai *V* minimum oleh pengelompokan *K-Means* dengan *k*=7, yaitu penentuan nilai *k* optimal menggunakan metode *elbow*. *Cluster* yang terbentuk memiliki karakteristik yang berbeda-beda sehingga dapat dijadikan

acuan dalam menentukan kegiatan yang diinginkan sesuai kebutuhan desa dalam *cluster*. Rincian karakteristik setiap *cluster* dapat dilihat melalui kolom keterangan pada tabel 6.

Tabel 6. Hasil *clustering*

No	Desa	Keterangan
1	Jatimulya	Memiliki jumlah KK yang relatif tinggi, KK yang terdampak tinggi, jarang terjadi banjir namun genangan yang terjadi dapat tinggi
2	Kudaile, Tembok Banjaran	Memiliki luas lahan yang rendah, kepadatan tinggi dan paling sering terjadi banjir
3	Padaharja, Jatilawang, Kemantran, Kemuning, Kramat, Dampyak, Dukuhturi, Kalisapu, Pecabean, Pedeslohor, Bersole, Tembok Lor, kaliwadas, Prupuk Selatan, Wotgalih, Blubuk, Sokareja, Banjaragung, Sidamulya	Memiliki kepadatan sedang, lahan sedang, kadang terjadi dengan genangan dari rendah menuju sedang ke tinggi.
4	Kupu, Lemah Duwur, Pagiyanten, Pesarean	Memiliki kepadatan dan genangan yang relatif tinggi namun jarang terjadi banjir
5	Maribaya, Mejasem, Sidaharja, Sidakaton, Sidapurna	Memiliki lahan sawah dan bukan sawah yang hampir seimbang, jumlah penduduk yang paling tinggi dan kk yang terdampak relatif tinggi
6	Semedo	Memiliki kepadatan dan jumlah kk paling rendah

Hasil *withinss* (*within sum of square*) dan *Betweenss* (*between sum of square*) dari penerapan *K-Means* menggunakan R selain menunjukkan jarak antar anggota dan antar *cluster* [11], dapat juga dilakukan perhitungan manual untuk menghasilkan *variance within* dan *variance between*. Sesuai perbandingan *variance within* dan *variance between* dihasilkan nilai *V* yang menunjukkan bahwa dengan metode *elbow* dihasilkan jumlah *cluster* yang ideal menggunakan algoritme *K-means*. Nilai *V* dari *clustering K-Means* menggunakan metode *elbow* lebih minimum dibandingkan penentuan *k* optimal menggunakan metode *silhouette* sesuai dengan definisi dari [6] mengenai *cluster ideal*. Penentuan nilai *k* optimal menggunakan *elbow* memberikan hasil 7 *cluster* dengan karakteristik yang berbeda antar kelompok namun sama antar objek dalam *cluster*. Setiap *cluster* memiliki karakteristik sesuai dengan faktor pemicu banjir pada [15], seperti *cluster* 2 yang memiliki kepadatan penduduk yang tinggi sehingga akan sering terjadi banjir dan *cluster* 7 yang kurang memiliki lahan resapan hingga saat curah hujan tinggi air yang akan tergenang juga tinggi.

#### 4. KESIMPULAN

Dari penerapan *K-Means* pada data rekapan laporan kejadian banjir Bidang I Kesiapsiagaan Bencana BPBD Kabupaten Tegal dan pengujian dalam menentukan nilai *k* optimal menggunakan metode *elbow* dan *silhouette*. Maka diperoleh kesimpulan bahwa hasil perbandingan nilai *V* menunjukkan penerapan metode *elbow* pada algoritme *K-Means* lebih ideal. *Cluster* yang terbentuk yaitu, *cluster* 1 : desa Jatimulya, *cluster* 2 : desa Kudaile dan Tembok Banjaran, *cluster* 3 : desa Padaharja, Jatilawang, Kemantran, Kemuning, Kramat, Dampyak, Dukuhturi, Kalisapu, Pecabean, Pedeslohor, Bersole, Tembok Lor, kaliwadas, Prupuk Selatan, Wotgalih, Blubuk, Sokareja, Banjaragung, Sidamulya, *cluster* 4 : desa Kupu, Lemah Duwur, Pagiyanten, Pesarean, *cluster* 5 : desa Maribaya, Mejasem, Sidaharja, Sidakaton, Sidapurna, *cluster* 6: Semedo, dan *cluster* 7 : Kedungsukun. Nilai *tot.withinss* sebesar 66.83625 yang lebih kecil dan *betweenss* yang lebih besar yaitu 189.1637 menunjukkan jarak dekat antar objek per *cluster* dan jarak yang jauh antar *cluster*

## 5. SARAN

Saran yang dapat diperhatikan untuk penelitian selanjutnya ialah memaksimalkan pengelompokan dengan tambahan data yang berpengaruh langsung pada bencana banjir, seperti mengenai curah hujan, DAS dan pemukiman sekitar bantaran sungai. Saran berikutnya untuk pengembangan lebih baik perlu dilakukan penerapan metode lain untuk pengelompokan atau untuk memaksimalkan setiap tahapan pengelompokan. Serta melakukan perbandingan untuk menentukan pengelompokan yang lebih spesifik dalam menghasilkan karakteristik.

## DAFTAR PUSTAKA

- [1] M. G. Sadewo, A. P. Windarto and D. Hartama, 2019, Algoritma K-Means Dalam Mengelompokkan Desa/Kelurahan Menurut Keberadaan Keluarga Pengguna Listrik dan Sumber Penerangan Jalan Utama Berdasarkan Provinsi, *Seminar nasional teknologi Komputer & Sains (SAINTEKS)*, pp. 754-761, Januari.
- [2] G. Shmueli, P. C. Bruce, I. Yahav, N. R. Patel and K. C. Lichtendahl Jr., 2018, *Data Mining for Business Analytics Concepts, Techniques, and Applications in R*, John Wiley & Sons, Inc.
- [3] L. Listiani, Y. H. Agustin and M. Z. Ramdhani, 2019, Implementasi Algoritma K-Means Cluster Untuk Rekomendasi Pekerjaan Berdasarkan Pengelompokan Data Penduduk, In *Sensitif 2019*, Tasikmalaya.
- [4] T. Ariwibowo, 2019, Perbandingan Metode Imputasi Mean, Median, Modus, Dan 1-Nn Pada Hasil Klasifikasi K-Nearest Neighbour (K-Nn) Studi Kasus : Klasifikasi Penyakit Jantung Koroner, Universitas Pembangunan Nasional Veteran.
- [5] H. Satriawan, 2018, Problematika Pembelajaran Matematika Pada Materi Statistika Smp Kelas IX, *Jurnal Elektronik Pembelajaran Matematika*, vol. 5, pp. 278-285, Oktober.
- [6] N. Asiska, N. Satyahadewi and H. Perdana, 2019, Pencarian Cluster Optimum Pada Single Linkage, Complete Linkage Dan Average Linkage, *Buletin Ilmiah Math, Stat, dan Terapannya (Bimaster)*, vol. 8, pp. 393-398.
- [7] R. I. Fajriah, H. Sutisna and B. K. Simpony, 2019, Perbandingan Distance Space Manhattan Dengan Euclidean Pada K-Means Clustering Dalam Menentukan Promosi, *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 4, pp. 36-49, Mei.
- [8] A. F. Febrianti, A. H. Cabral and G. Anuraga, 2018, K-Means Clustering Dengan Metode Elbow Untuk Pengelompokan Kabupaten Dan Kota Di Jawa Timur Berdasarkan Indikator Kemiskinan, in *SNHRP-1*.
- [9] C. D. Rumiati and I. Budi, 2017, Segmentasi Pelanggan Pada Customer Relationship Management Di Perusahaan Ritel: Studi Kasus Pt Gramedia Asri Media, *Jurnal Sistem Informasi (Journal of Information Systems)*, vol. 13, no. 1, pp. 1-10, April.
- [10] B. Santoso, I. Cholissodin and B. D. Setiawan, "Optimasi K-Means untuk Clustering Kinerja Akademik Dosen Menggunakan Algoritme Genetika," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, pp. 1652-1659, 2017.
- [11] Y. Novianti, 2017, Implementasi K-Means Clustering Algorithm Untuk Analisa Soal Ujian Online Pada Smp Islam Al-Fath Pare, *Simki-Techsain*, vol. 1.
- [12] Kusnendi, 2016, Memahami Analisis Varians.
- [13] K-Means Clustering, <https://stat.ethz.ch/R-manual/R-patched/library/stats/html/k-means.html>, diakses tgl 20 Februari 2020.

- [14] R. Adrianto and A. Fahmi, 2016, Penerapan Metode Clustering Dengan Algoritma K-Means Untuk Rekomendasi Pemilihan Jalur Peminatan Sesuai Kemampuan Pada Program Studi Teknik Informatika - S1 Universitas Dian Nuswantoro, pp. 101-116.
  - [15] E. Yulaelawati, Ph.d and U. Syihab, Ph.d, 2008, Mencerdasi Bencana : banjir, tanah longsor, tsunami, gempa bumi, gunung api, kebakaran, Grasindo.
-