

## Penerapan Metode *Simple K-Medoid* (SKM) dalam Mengklasterkan SMA di Kabupaten Karanganyar Tahun 2023

Dwi Ayu Puspita Sari\*, Nur Azizah Komara Rifai

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

\*dwiayupuspita13@gmail.com, nur.azizah@unisba.ac.id

**Abstract.** Cluster analysis is an analysis that classifies objects that have similar characteristics into the same group. One of the most common cluster analysis methods is the non-hierarchical method where the number of clusters is determined by the researcher. Some of the methods included in non-hierarchical methods are k-means and k-medoid. The k-means method is a clustering method used for numeric variable data, while the k-medoid method can be used for mixed variable data (numeric and categorical data). The k-medoid method used for mixed variable data is the Simple K-Medoid (SKM) method. The cluster results obtained in the SKM method must be validated using the Medoid Shadow Value (MSV) to see the best number of clusters. The data used in this study comes from the Sekolah Kita website and the Basic Education Data which is mixed data consisting of 14 numerical variables and 3 binary and categorical variables. Based on the results of the SKM method clustering and cluster validation using MSV, it was found that there were 2 clusters of high schools in Karanganyar District. Cluster 1 consists of 15 high schools from 18 high schools with more complete facilities and infrastructure than cluster 2 which consists of 3 high schools from 18 high schools.

**Keywords:** *Cluster Analysis, Medoid Shadow Value, Simple K-Medoid.*

**Abstrak.** Analisis kluster adalah analisis yang mengklasifikasikan objek yang memiliki kesamaan karakteristik ke dalam kelompok yang sama. Salah satu metode analisis kluster yang paling umum adalah metode non hierarki di mana jumlah kluster ditentukan oleh peneliti. Beberapa metode yang termasuk ke dalam metode non hierarki adalah *k-means* dan *k-medoid*. Metode *k-means* adalah metode kluster yang digunakan untuk data bervariabel numerik, sedangkan metode *k-medoid* dapat digunakan untuk data bervariabel campuran (data numerik dan kategorik). Metode *k-medoid* yang digunakan untuk data bervariabel campuran adalah metode *Simple K-Medoid* (SKM). Hasil kluster yang didapat pada metode SKM harus divalidasi menggunakan *Medoid Shadow Value* (MSV) untuk melihat jumlah kluster yang terbaik. Data yang digunakan pada penelitian ini berasal dari website Sekolah Kita dan Data Pokok Pendidikan yang merupakan data campuran yakni terdiri dari 14 variabel numerik dan 3 variabel biner dan kategorik. Berdasarkan hasil dari pengelompokan metode SKM dan validasi kluster menggunakan MSV didapatkan pengelompokan SMA di Kabupaten Karanganyar sebanyak 2 kluster. Kluster 1 beranggotakan 15 SMA dari 18 SMA dengan memiliki sarana dan prasarana yang lebih lengkap dibandingkan kluster 2 yang beranggotakan 3 SMA dari 18 SMA.

**Kata Kunci:** *Analisis Kluster, Medoid Shadow Value, Simple K-Medoid.*

## A. Pendahuluan

Analisis kluster adalah analisis yang mengklasifikasikan objek yang memiliki kesamaan karakteristik ke dalam kelompok yang sama. Salah satu metode analisis kluster yang paling umum adalah metode non hierarki. Metode non hierarki bertujuan untuk mengelompokkan  $n$  objek ke dalam kelompok [1]. Beberapa metode non hierarki yang banyak digunakan yaitu metode  $k$ -means dan  $k$ -medoid [2]

Metode  $k$ -medoid diciptakan untuk mengatasi kelemahan yang ada pada  $k$ -means yang sensitif pada data pencilan. Metode  $k$ -means adalah metode kluster yang digunakan pada data numerik. Namun, pada kenyataannya data tidak selalu bervariasi numerik terkadang data juga dapat bervariasi campuran (data numerik dan kategorik), untuk mengatasi hal tersebut maka metode kluster yang dapat digunakan adalah metode  $k$ -medoid. Salah satu metode  $k$ -medoid yang dapat digunakan pada data campuran adalah metode *Simple K-Medoid* (SKM) [3]. Data yang terdiri dari data campuran salah satunya adalah data terkait pendidikan.

Pendidikan merupakan salah satu kegiatan universal untuk mempersiapkan dan mengembangkan sumber daya manusia sehingga, mampu menghadapi tantangan dan memiliki masa depan yang lebih baik [4]. Salah satu kualifikasi pendidikan yang harus diselesaikan adalah Sekolah Menengah Atas (SMA). SMA merupakan salah satu jenjang sekolah yang harus diperhatikan pemilihannya, terutama pada sarana dan prasarana yang dimiliki suatu sekolah. Sarana dan prasarana yang mendukung akan menunjang proses pembelajaran menjadi lebih efektif. Menurut Kahfi dan Widiyastuti (2020) sarana dan prasarana yang terdapat pada SMA di Kabupaten Karanganyar masih kurang merata. Untuk mempermudah pemerintah dalam penanganan pemerataan kebutuhan sekolah seperti sarana dan prasarana, maka diperlukan pengklusteran atau pengelompokan sekolah [6].

Menurut penelitian yang dilakukan oleh Desita (2023), kualitas suatu SMA dapat ditentukan dengan melihat sarana dan prasarana di sekolah tersebut. Sarana dan prasarana sekolah dapat terdiri dari data campuran sehingga, klusterisasi SMA dapat menggunakan metode SKM. Hasil kluster yang didapat harus divalidasi untuk melihat jumlah kluster yang terbaik [8]. Metode validasi kluster yang cocok untuk metode SKM yaitu metode *Medoid Shadow Value* (MSV) [9]. Maka dari itu berdasarkan penjelasan di atas, penulis tertarik untuk menerapkan metode *Simple K-Medoid* (SKM) dalam mengklusterkan SMA di Kabupaten Karanganyar tahun 2023. Selanjutnya, tujuan dalam penelitian ini diuraikan sebagai berikut:

1. Mengelompokkan SMA di Kabupaten Karanganyar menggunakan metode *Simple K-Medoid* (SKM) berdasarkan sarana dan prasarana sekolah.
2. Melakukan validasi kluster dengan menggunakan metode *Medoid Shadow Value* (MSV).

## B. Metodologi Penelitian

Dalam penelitian ini data yang digunakan adalah data sekunder, yaitu data SMA di Kabupaten Karanganyar yang diperoleh dari website Sekolah Kita dan Data Pokok Pendidikan sebanyak 18 sekolah. Adapun variabel pada penelitian yang digunakan adalah status sekolah, sumber listrik, akses internet, ruang kelas, perpustakaan, laboratorium, ruang pimpinan, ruang guru, ruang ibadah, usaha kesehatan sekolah (UKS), toilet, gudang, sirkulasi, tempat olahraga, ruang tata usaha, ruang konseling, ruang organisasi siswa intra sekolah (OSIS). Analisis data pada penelitian ini menggunakan software R Studio dan Microsoft Excel.

### Langkah Analisis Data

Mencari hasil kluster menggunakan metode *Simple K-Medoid* (SKM) dan melakukan validasi hasil kluster menggunakan metode *Medoid Shadow Value* (MSV) sebagai berikut:

1. Menghitung jarak objek menggunakan jarak Gower, adapun perhitungan jarak Gower sebagai berikut:
  - a. Menghitung nilai  $s_{ij}$  dengan menggunakan persamaan sebagai berikut:

$$s_{ij} = \frac{\sum_{l=1}^p w_{ijl} s_{ijl}}{\sum_{l=1}^p w_{ijl}} \quad (1)$$

Jika, variabel numerik maka  $s_{ijl}$  dihitung menggunakan persamaan sebagai

berikut:

$$s_{ijl} = 1 - \frac{|x_{il} - x_{jl}|}{R_l}, l = 1, 2, \dots, p \quad (2)$$

Di mana:

$s_{ijl}$ : Kemiripan antara objek ke- $i$  dan ke- $j$  variabel numerik ke- $l$ .

$x_{il}$ : Nilai dari objek ke- $i$  pada variabel numerik ke- $l$ .

$x_{jl}$ : Nilai dari objek ke- $j$  pada variabel numerik ke- $l$ .

$R_l$ : Rentang atau Jangkauan dari variabel numerik ke- $l$ .

dan jika, variabelnya biner dan kategorik maka  $s_{ijl}$  sebagai berikut:

$$s_{ijl} \in \{0, 1\} \quad (3)$$

Di mana:

$s_{ijl} : 1$ , berarti bahwa objek ke- $i$  dan ke- $j$  sama pada variabel ke- $l$ .

$s_{ijl} : 0$ , berarti bahwa objek ke- $i$  dan ke- $j$  berbeda pada variabel ke- $l$ .

- b. Menghitung nilai  $d_{ij}$  dengan menggunakan persamaan sebagai berikut:

$$d_{ij} = 1 - s_{ij} \cdot w_{ijl} \quad (4)$$

Di mana jika pada data terdapat *missing value* maka nilai dari  $w_{ijl}$  adalah 0.

Perhitungan pada jarak Gower mengasumsikan bahwasannya tidak ada *missing value* dan setiap variabel memiliki kontribusi yang sama sehingga untuk  $w_{ijl}$  bernilai sama dengan 1.

2. Menentukan jumlah  $k$  kluster.
3. Menentukan pusat kluster (medoid) awal dengan menghitung nilai  $a_i$  dengan menggunakan persamaan sebagai berikut:

$$a_i = \sum_{j=1}^n d_{ij} = \sum_{j=1}^n d_{ji}, i = 1, 2, 3, \dots, n \quad (5)$$

Di mana  $a_i$  merupakan jumlah jarak dari objek  $i$  ke objek lainnya. Selanjutnya memilih nilai pusat kluster dengan melihat nilai minimum pada nilai  $a_i$  untuk pusat kluster pertama dan pusat kluster selanjutnya dipilih secara *random*.

4. Menghitung jarak antar objek ke medoid awal dan tentukan jarak terdekatnya. Menghitung total jarak terdekat ( $E$ ) objek ke medoid awal, dengan menggunakan persamaan sebagai berikut:

$$E = \sum_{p=1}^k \sum_{n:l(x_n)=m_p} d(x_n, m_p) \quad (6)$$

Selanjutnya tetapkan label keanggotaan kluster untuk setiap objek.

5. Menentukan medoid baru dengan mempertahankan label kluster. Penentuan medoid baru menggunakan persamaan sebagai berikut:

$$m_p := \underset{p \in K}{\operatorname{argmin}} \sum_{n:l(x_n)=m_p} d(x_n, m_p), p = 1, 2, 3, \dots, k \quad (7)$$

Di mana:

$m_p$ : Sebuah objek yang memiliki jumlah jarak yang paling minimum dalam satu kluster.

6. Menghitung jarak antar objek ke medoid baru dan tentukan jarak terdekatnya. Menghitung total jarak terdekat ( $E$ ) objek ke medoid baru, dengan menggunakan persamaan (6), lalu tetapkan label keanggotaan kluster untuk setiap objek.
7. Ulangi langkah 5 dan 6 untuk mendapatkan nilai total jarak terdekat ( $E$ ) yang baru sama dengan total jarak terdekat ( $E$ ) lama.
8. Ulangi langkah 3 sampai 7, untuk mendapatkan beberapa hasil *seeding* secara acak dari medoid awal. Objek yang selalu berada di tengah akan selalu menjadi salah satu medoid awal. Penentuan medoid akhir dilihat dari nilai total jarak terdekat ( $E$ ) yang paling minimum.
9. Tetapkan keanggotaan kluster untuk setiap objek ke medoid akhir.
10. Lakukan pengecekan kluster menggunakan metode validasi *Medoid Shadow Value* (MSV) dengan menggunakan persamaan sebagai berikut:

$$msv(x) = \frac{d(x, m^{\prime}(x)) - d(x, m(x))}{d(x, m^{\prime}(x))} \tag{8}$$

Di mana:

$d(x, m(x))$ : Jarak antara objek  $x$  ke medoid terdekat pertama.

$d(x, m^{\prime}(x))$ : Jarak antara objek  $x$  ke medoid terdekat kedua.

11. Menghitung MSV global menggunakan persamaan sebagai berikut:

$$msv\ global = \frac{1}{n} \sum_{x=1}^n msv(x) \tag{9}$$

Di mana:

$msv(x)$  : Nilai MSV dari data (objek) ke-  $x$ .

$n$  : Banyaknya data.

Langkah terakhir yaitu melakukan profilisasi hasil kluster untuk menggambarkan karakteristik masing-masing kluster dan dapat ditarik kesimpulan dari penelitian ini.

### C. Hasil Penelitian dan Pembahasan

#### Pengukuran Jarak Gower Antar Objek Variabel Campuran

Tujuan dari analisis kluster adalah mengelompokkan objek-objek yang mirip menjadi satu kluster. Untuk mengetahui seberapa mirip antar objek maka dilakukan pengukuran kesamaan, salah satunya menggunakan pengukuran kesamaan berdasarkan jarak (*distance*). Adapun matriks jarak antar objek disajikan pada tabel berikut:

**Tabel 1.** Matriks Jarak Antar Objek

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	...	<b>18</b>
<b>1</b>	0,0000	0,1287	0,2537	0,2517	0,3149	0,3894	...	0,3668
<b>2</b>	0,1287	0,0000	0,2789	0,2392	0,2813	0,2970	...	0,2970
<b>3</b>	0,2537	0,2789	0,0000	0,1522	0,1741	0,1819	...	0,4018
<b>4</b>	0,2517	0,2392	0,1522	0,0000	0,2526	0,2382	...	0,3362
<b>5</b>	0,3149	0,2813	0,1741	0,2526	0,0000	0,2824	...	0,4630
<b>6</b>	0,3894	0,2970	0,1819	0,2382	0,2824	0,0000	...	0,4552
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<b>18</b>	0,3668	0,2970	0,4018	0,3362	0,4630	0,4552	...	0,0000

#### Penentuan Pusat Kluster (Medoid) Awal

Pemilihan pusat kluster (medoid) awal ditentukan dengan melihat jumlah jarak ( $a_i$ ) yang paling minimal. Adapun nilai  $a_i$  sebagai berikut:

**Tabel 2.** Nilai  $a_i$

<b>Data</b>	<b><math>a_i</math></b>	<b>Data</b>	<b><math>a_i</math></b>	<b>Data</b>	<b><math>a_i</math></b>
1	4,8330	7	4,7084	13	3,9383
2	4,6351	8	7,5193	14	4,3652
3	4,1540	9	4,7308	15	5,6315
4	4,3739	10	5,4873	16	4,9021
5	5,4605	11	4,1187	17	6,3184
6	5,3446	12	7,0846	18	6,6683

Berdasarkan tabel diatas dapat dilihat bahwa data ke-13 memiliki nilai  $a_i$  terkecil sebesar 3,9383, yang berarti bahwa pusat kluster (medoid) awal adalah data ke-13 untuk kluster 1. Untuk menentukan pusat kluster (medoid) awal untuk kluster selanjutnya dipilih secara *random* dari data.

**Pengelompokkan Menggunakan *Simple K-Medoids***

Pada penelitian ini, pengelompokkan SMA di Kabupaten Karanganyar akan di klusterkan menjadi 2 kelompok dan 3 kelompok. Pengelompokkan dilakukan dengan *seeding* sebanyak 20 kali dan iterasi sebanyak 10 kali. Adapun hasil ringkasan proses pengelompokkan untuk 2 kelompok disajikan pada tabel berikut:

**Tabel 3.** Hasil Perhitungan dengan Seeding 20 Untuk 2 Klaster

Seeding	Iterasi	Medoid		Nilai E	Seeding	Iterasi	Medoid		Nilai E
		M1	M2				M1	M2	
1	1	13	4	34,611	9	1	13	17	29,551
	2	13	3	29,404		2	13	8	29,050
	3	13	3	29,404		3	13	8	29,050
2	1	13	7	37,012	10	1	13	6	33,160
	2	13	3	29,404		2	13	3	29,404
	3	13	3	29,404		3	13	3	29,404
3	1	13	11	29,522	11	1	13	7	37,012
	2	13	3	29,404		2	13	3	29,404
	3	13	3	29,404		3	13	3	29,404
4	1	13	3	29,404	12	1	13	16	37,437
	2	13	3	29,404		2	13	15	37,057
5	1	13	5	35,244		3	13	15	37,057
	2	13	3	29,404	13	1	13	18	36,979
	3	13	3	29,404		2	13	18	36,979
6	1	13	17	29,551	14	1	13	2	37,415
	2	13	8	29,050		2	3	13	29,404
7	1	13	11	29,522		3	3	13	29,404
	2	13	3	29,404	15	1	13	16	37,437
	3	13	3	29,404		2	13	15	37,057
8	1	13	6	33,160	16	3	13	15	37,057
	2	13	3	29,404		1	13	17	29,551
	3	13	3	29,404		2	13	8	29,050

Seeding	Iterasi	Medoid		Nilai E	Seeding	Iterasi	Medoid		Nilai E
		M1	M2				M1	M2	
16	3	13	8	29,050	19	1	13	3	29,404
17	1	13	2	37,415		2	13	3	29,404
	2	3	13	29,404	20	1	13	6	33,160
	3	3	13	29,404		2	13	3	29,404
18	1	13	2	37,415	3	13	3	29,404	
	2	3	13	29,404					
	3	3	13	29,404					

Berdasarkan tabel diatas didapatkan hasil dari 20 kali *seeding* dengan diperoleh nilai *E* yang paling minimum sebesar 2,9050. Nilai *E* tersebut dihasilkan dari medoid akhir yaitu objek ke-13 dan objek ke-8. Maka pengelompokkan menunjukkan bahwa klaster 1 terdiri dari 15 sekolah dan klaster 2 terdiri dari 3 sekolah. Selanjutnya hasil ringkasan proses pengelompokkan untuk 3 kelompok disajikan pada tabel berikut:

**Tabel 4.** Hasil Perhitungan dengan Seeding 20 Untuk 3 Klaster

Seeding	Iterasi	Medoid			Nilai E	Seeding	Iterasi	Medoid			Nilai E		
		M1	M2	M3				M1	M2	M3			
1	1	13	4	7	32,851	9	3	3	13	9	27,549		
	2	13	3	6	26,934			10	1	13	17	2	35,315
	3	13	3	6	26,934				2	13	8	2	27,081
2	1	13	11	3	28,934	11	3	13	8	2	27,081		
	2	13	8	3	22,127			1	13	2	3	27,459	
	3	13	8	3	22,127			2	13	2	3	27,459	
3	1	13	5	17	26,106	12	1	13	6	17	26,021		
	2	13	3	8	22,127			2	13	3	8	22,127	
	3	13	3	8	22,127			3	13	3	8	22,127	
4	1	13	11	2	27,576	13	1	13	10	18	27,421		
	2	13	3	2	27,459			2	3	13	18	27,000	
	3	13	3	2	27,459			3	3	13	18	27,000	
5	1	13	6	9	31,305	14	1	13	11	4	28,284		
	2	13	3	9	27,549			2	13	3	4	28,166	
	3	13	3	9	27,549			3	13	3	4	28,166	
6	1	13	17	6	26,021	15	1	13	14	9	30,587		
	2	13	8	3	22,127			2	13	3	9	27,549	
	3	13	8	3	22,127			3	13	3	9	27,549	
7	1	13	7	16	35,066	16	1	13	6	8	25,520		
	2	13	3	15	27,078			2	13	3	8	22,127	
	3	13	3	15	27,078			3	13	3	8	22,127	
8	1	13	18	2	35,011	17	1	13	9	18	35,124		
	2	3	18	13	27,000			2	3	13	18	27,000	
	3	3	18	13	27,000			3	3	13	18	27,000	
9	1	13	2	16	35,537	18	1	13	10	16	35,808		
	2	3	13	9	27,549			2	3	13	15	27,078	

Seeding	Iterasi	Medoid			Nilai E	Seeding	Iterasi	Medoid			Nilai E	
		M1	M2	M3				M1	M2	M3		
18	3	3	13	15	27,078	20	1	13	1	10	35,293	
19	1	13	9	16	35,967			2	3	13	10	27,775
	2	3	13	1	27,220			3	3	13	10	27,775
	3	3	13	1	27,220							

Berdasarkan tabel diatas didapatkan hasil dari 20 kali *seeding* dengan diperoleh nilai *E* yang paling minimum sebesar 2,2127. Nilai *E* tersebut dihasilkan dari medoid akhir yaitu objek ke-13, objek ke-3 dan objek ke-8. Maka pengelompokkan menunjukkan bahwa klaster 1 terdiri dari 10 sekolah, klaster 2 terdiri dari 5 sekolah dan klaster 3 terdiri dari 3 sekolah.

#### Validasi Klaster Menggunakan *Medoid Shadow Value* (MSV)

Hasil pengelompokkan menggunakan metode SKM divalidasi menggunakan metode *Medoid Shadow Value* (MSV). Validasi dilakukan dengan menggunakan percobaan  $k=2$  dan  $k=3$ . Adapun penjelasan metode MSV dengan percobaan  $k=2$  untuk objek 1 sebagai berikut:

$$msv(x) = \frac{d(x, m'(x)) - d(x, m(x))}{d(x, m'(x))} = msv(1) = \frac{0,4659 - 0,1626}{0,4659} = 0,65102$$

Perhitungan dilakukan untuk seluruh data (objek), yang disajikan pada tabel berikut:

**Tabel 5.** Nilai MSV: Percobaan dengan 2 Klaster

Data	Nilai MSV	Klaster
1	0,6510	1
2	0,7655	1
⋮	⋮	⋮
17	0,6188	2
18	0,6427	1

Penjelasan metode MSV dengan percobaan  $k=3$  untuk objek 1 sebagai berikut:

$$msv(x) = \frac{d(x, m'(x)) - d(x, m(x))}{d(x, m'(x))} = msv(1) = \frac{0,2537 - 0,1626}{0,2537} = 0,3591$$

Perhitungan dilakukan untuk seluruh data (objek), yang disajikan pada tabel berikut:

**Tabel 6.** Nilai MSV: Percobaan dengan 3 Klaster

Data	Nilai MSV	Klaster
1	0,3591	1
2	0,5376	1
⋮	⋮	⋮
17	0,5732	3
18	0,4017	1

Penentuan jumlah klaster terbaik dengan perhitungan sebagai berikut:

- MSV global untuk jumlah klaster 2

$$msv\ global = \frac{1}{n} \sum_{x=1}^n msv(x)$$

$$msv\ global = \frac{0,6510 + 0,7655 + 0,3272 + \dots + 0,6188 + 0,6427}{18} = 0,6374$$

- MSV global untuk jumlah klaster 3

$$msv\ global = \frac{1}{n} \sum_{x=1}^n msv(x)$$

$$msv\ global = \frac{0,3591 + 0,5376 + 1,0000 + \dots + 0,5732 + 0,4017}{18} = 0,5831$$

Penentuan jumlah klaster terbaik ditentukan apabila nilai  $msv$  global mendekati 1. Jika dilihat dari perhitungan diatas, nilai  $msv$  global tertinggi antara 2 percobaan jumlah klaster adalah jumlah klaster sebanyak 2 dengan nilai  $msv$  global sebesar 0,6374, sehingga dapat dikatakan bahwa jumlah  $k$  klaster yang terbaik untuk pengelompokkan SMA di Kabupaten Karanganyar berdasarkan sarana dan prasana dengan menggunakan metode SKM adalah 2 klaster.

### Hasil Klaster

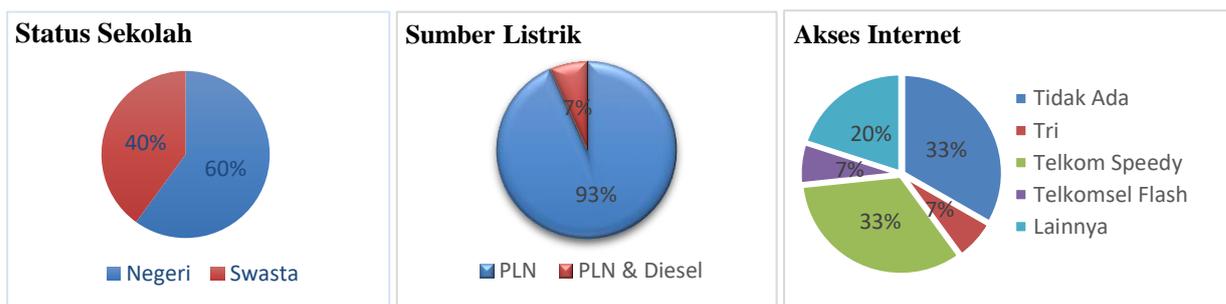
Berdasarkan hasil dari validasi klaster, jumlah klaster yang digunakan berjumlah  $k=2$  sehingga pengelompokkan yang terbentuk disajikan pada tabel berikut:

**Tabel 7.** Hasil Pengelompokan 2 Klaster

Klaster	Jumlah Anggota	Nama Sekolah
1	15	SMAN 1 Karanganyar, SMAN 2 Karanganyar, SMA Kanisius Bharata Karanganyar, SMA Muhammadiyah 1 Karanganyar, SMA Muhammadiyah Daruk Arqom, SMA Muhammadiyah 5 Jaten, SMAN Colomadu, SMAN Mojogedang, SMAN Karangpandan, SMA Bung Karno Karangpandan, SMAN Gondangrejo, SMA Muhammadiyah 3 Gondangrejo, SMAN Jumapolo, SMAN Kebakkramat, SMAN Kerjo.
2	3	SMA Sains Daarul Quran Karanganyar, SMA Islam Al-Azhar 28 IIBS Karanganyar, SMAN Tawangmangu.

**Profilisasi Hasil Klaster**

Profilisasi hasil klaster diperlukan untuk mengetahui karakteristik dari masing-masing klaster. Pada penelitian ini digunakan data campuran sehingga terdapat dua karakteristik yaitu data numerik dan kategorik. Karakteristik klaster pada data kategorik disajikan dalam diagram lingkaran berikut:



**Gambar 1.** Karakteristik Kategorik Untuk Klaster 1



**Gambar 2.** Karakteristik Kategorik Untuk Klaster 2

Berdasarkan Gambar 1 dan Gambar 2 diatas diperoleh informasi karakteristik klaster untuk data kategorik sebagai berikut:

Klaster 1: pada klaster 1 di dominasi dengan sekolah SMA berstatus negeri dengan sumber listrik PLN dan kebanyakan sekolah sudah memiliki akses internet.

Klaster 2: pada klaster 2 di dominasi dengan sekolah SMA berstatus swasta dengan sumber listrik PLN dan kebanyakan tidak memiliki akses internet.

Karakteristik klaster pada data numerik dilihat dari rata-rata disajikan dalam tabel

berikut:

**Tabel 8.** Profilisasi Anggota Klaster Berdasarkan Variabel Numerik

Variabel	Klaster		Variabel	Klaster	
	1	2		1	2
X <sub>4</sub>	21,7333	2,3333	X <sub>11</sub>	9,6	1
X <sub>5</sub>	1,1333	0,6667	X <sub>12</sub>	1,2	0
X <sub>6</sub>	5,1333	1,3333	X <sub>13</sub>	0,1333	0
X <sub>7</sub>	0,9333	1	X <sub>14</sub>	0,6	0
X <sub>8</sub>	1,0667	0,3333	X <sub>15</sub>	1	0
X <sub>9</sub>	1,2667	0	X <sub>16</sub>	1,4667	0
X <sub>10</sub>	1,0667	0	X <sub>17</sub>	0,9333	0

Berdasarkan Tabel 8 diatas diperoleh informasi karakteristik klaster untuk data numerik sebagai berikut:

Klaster 1: pada klaster 1, jika dilihat dari rata-rata jumlah ruang kelas, perpustakaan, laboratorium, ruang pimpinan, ruang guru, ruang ibadah, UKS, toilet, gudang, sirkulasi, tempat olahraga, ruang TU, ruang konseling, ruang OSIS, klaster 1 memiliki rata-rata yang lebih tinggi dibandingkan dengan klaster 2, yang berarti klaster 1 memiliki sarana dan prasarana yang lebih lengkap dibandingkan klaster 2.

Klaster 2: pada klaster 2, jika dilihat dari rata-rata jumlah ruang kelas, perpustakaan, laboratorium, ruang pimpinan, ruang guru, ruang ibadah, UKS, toilet, gudang, sirkulasi, tempat olahraga, ruang TU, ruang konseling, ruang OSIS, klaster 2 memiliki rata-rata yang lebih rendah dibandingkan dengan klaster 1, yang berarti klaster 2 memiliki sarana dan prasarana yang kurang lengkap dibandingkan klaster 1.

#### D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

1. Pengklasteran SMA di Kabupaten Karanganyar menggunakan metode *Simple K-Medoid* (SKM) menghasilkan 2 klaster. Klaster 1 beranggotakan 15 SMA dari 18 SMA dengan sarana dan prasarana yang lebih lengkap dibandingkan klaster 2 yang beranggotakan 3 SMA dari 18 SMA.
2. Berdasarkan hasil perhitungan *Medoid Shadow Value* (MSV) didapatkan jumlah *k* klaster yang terbaik pada pengelompokan SMA di Kabupaten Karanganyar berdasarkan sarana dan prasarana adalah 2 klaster dengan nilai MSV global sebesar 0,6374.

#### Acknowledge

Penulis mengucapkan terima kasih kepada Allah SWT, kedua orang tua yang memberikan dukungan, para dosen Statistika Unisba yang telah memberikan ilmu pengetahuannya, teman-teman yang telah membantu dan memberi semangat kepada penulis sehingga dapat menyelesaikan penelitian ini.

#### Daftar Pustaka

- [1] S. M. Aziz and N. A. K. Rifai, "Pengelompokan Ekspor Kopi Menurut Negara Tujuan Menggunakan Metode K-Means Clustering dengan Silhouette Coefficient," Bandung Conference Series: Statistics, vol. 2, no. 2, p. 417, Aug. 2022, doi: 10.29313/bcss.v2i2.4536.
- [2] I. Gunawan, Pengantar Statistika Inferensial. 2016.
- [3] W. Budiaji and F. Leisch, "Simple K-Medoids Partitioning Algorithm for Mixed Variable Data," Algorithms, vol. 12, no. 177, Sep. 2019, doi: 10.3390/a12090177.

- [4] Y. Alpian, S. W. Anggraeni, U. Wiharti, and N. M. Soleha, “Pentingnya Pendidikan bagi Manusia,” *Jurna Buana Pengabdian*, vol. 1, no. 1, 2019.
- [5] M. Al Kahfi and D. Widiyastuti, “Kajian Ketersediaan dan Pola Distribusi Fasilitas Pendidikan Sekolah Menengah Atas/ Sederajat di Kabupaten Karanganyar,” 2020.
- [6] Nurahman, A. Purwanto, and S. Mulyanto, “Klasterisasi Sekolah Menggunakan Algoritma K-Means Berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik,” *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 2, pp. 337–350, Mar. 2022, doi: 10.30812/matrik.v21i2.1411.
- [7] D. Desita, “Implementasi Metode Ensemble Robust Clustering Using Links (Rock) Untuk Klasterisasi Sekolah Menengah Atas (SMA) Di Bandar Lampung,” 2023.
- [8] T. Susilowati, D. Sugiarto, and I. Mardianto, “Uji Validasi Algoritma Self-Organizing Map (SOM) dan K-Means untuk Pengelompokan Pegawai,” *JURNAL RESTI*, vol. 1, no. 3, p. 1172, 2017.
- [9] W. Budiaji, “Medoid-based shadow value validation and visualization,” *International Journal of Advances in Intelligent Informatics*, vol. 5, no. 2, pp. 76–88, Jul. 2019, doi: 10.26555/ijain.v5i2.326.
- [10] Anggi Priliani Yulianto and S. Darwis, “Penerapan Metode K-Nearest Neighbors (kNN) pada Bearing,” *Jurnal Riset Statistika*, vol. 1, no. 1, pp. 10–18, Jul. 2021, doi: 10.29313/jrs.v1i1.16.
- [11] Salsabila Pratiwi and Marizsa Herlina, “Pengaruh Harga Pangan terhadap Inflasi dengan Metode Vector Autoregressive Integrated Moving Average,” *Jurnal Riset Statistika*, pp. 87–96, Dec. 2023, doi: 10.29313/jrs.v3i2.2690.
- [12] F. Dewi, H. 1□, and A. Kudus, “Penanganan Data Missing dengan Algoritma Multivariate Imputation By Chained Equations (MICE),” vol. 1, no. 1, pp. 35–42, 2023, doi: 10.29313/datamath.v1i1.25.